Faculty of Science, Technology, Engineering and Mathematics
M140 Introducing statistics

Unit 2 Prices

Contents

Introduction

This unit and Unit 3 examine, in various ways, the question:

Are people getting better or worse off?

Because this is a statistics module, we shall concentrate on the statistical aspects of the question. This unit focuses on statistics about prices, and Unit 3 moves on to consider statistics about earnings; this enables us to look at the question of whether earnings have been increasing more rapidly than prices.

However, it is not the case that statistics can provide all the answers – or even the best answer – to the question of whether people are getting better or worse off. There are many non-statistical issues which are relevant and it is important to put the statistical approach in its correct perspective. To take just one example: if earnings are rising rapidly but unemployment is also rising, then no statistical analysis based on a comparison of earnings with prices will have any relevance to the circumstances of a person who has become unemployed.

In the question examined in these units, *people* does not refer specifically to *you*, Open University students, but to the whole of society in the UK. That is quite a big batch (more than 62 million in 2010, according to an estimate from the UK's Office for National Statistics), consisting of men, women and children, living alone, in large or small households, or in institutions; some of them working, others unemployed, some retired and others not yet old enough for paid work.

It is not possible, using statistical techniques, to provide a complete answer to this one question covering such a big theme, particularly an answer which is valid for all these people and their varied economic and social circumstances; data and techniques both have to be used with common sense. Instead, the aim of these texts is more modest: to explore small batches of data relevant to the question (and relating to some individuals and groups in society), using basic analytical and graphical techniques.

See Unit 1 – Subsection 3.2, Section 4 and Section 5.

We start with price data and look at some different ways of measuring the overall *location* of a batch of price figures for a single item. In looking for patterns in data, the initial procedures are to round the figures, if necessary, in an appropriate and convenient way, then to draw a stemplot. The next step is to find a measure representing the location of the batch; this will be a value lying between the lowest and highest values of the batch. You have already met one important location measure: the median. (There will be more about this in what follows.) Another very important measure is the *arithmetic mean*, which is introduced in Subsection 1.3.

Section 2 shows how to calculate the *weighted mean*, which is a quantity related to the arithmetic mean. You will also learn about some circumstances where it makes sense to calculate a weighted mean.

Having considered the location of a batch, it is often helpful to examine the spread of values and the shape of the distribution of values between the extremes and around the average. Section 3 shows how to calculate one particular measure of spread for a batch: the *interquartile range*. It also shows some diagrammatic methods for representing the spread and shape of the distribution of values in a batch.

Section 4 introduces the notion of a *price index* for indicating changes in the price of a single item and for two or more different items. Section 5 looks at the UK's Retail Prices Index (RPI) and Consumer Prices Index (CPI), which measure changes in prices over time.

The central question, *Are people getting better or worse off?*, is partly addressed in this unit, which focuses on the 'prices' element. If prices are rising, then, other things being equal, we are worse off. It is left to Unit 3 to examine the other important element, 'earnings'. If our earnings are increasing, then, other things being equal, we are better off. However, other things are usually *not* equal – prices and earnings are generally changing at the same time, and Unit 3 also covers the question of how to deal with both sorts of changes at once.

Note that Section 5 is longer than all the other sections, so you should plan your study time accordingly.

Section 6 directs you to the Computer Book. You are also guided to the Computer Book after completing Section 1 and Subsection 2.1. It is better to do the work at those points in the text, although you can leave it until later if you prefer.

1 Measuring location

Measuring location has two components:

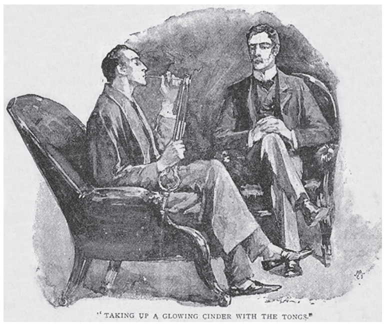
- gathering data about the quantity of interest
- determining a value to represent the location of the data.

The task of gathering appropriate data is somewhat problem-specific – general strategies are available, but exact details usually need to be decided for each problem. To determine the price of an electric kettle, for example, we would have to decide the size and type of kettle we're interested in, where and when its purchased, and so forth. In contrast, choosing a value to summarise the location of a set of data is more straightforward. In this section, we will focus on the two most common measures of location: the median and the mean. The data gathered about the quantity of interest does not affect the way we calculate these location measures.

1.1 Data on prices

In order to measure how prices change, we need data on prices and some way of measuring their overall location. Price data take many forms, some of which you have met in Unit 1.

In examining the overall location, prices of all goods are relevant, but some are more important than others. Ballpoint pens are relatively unimportant in most people's shopping baskets, coffee prices are unimportant for tea drinkers, and chicken prices are of little concern to vegetarians. Our first batch of price data is coffee prices (see Table 1).



'Data, data, data!' he cried impatiently. 'I can't make bricks without clay.' (Sherlock Holmes in *The Adventure of the Copper Beeches* by A.C. Doyle (1892))

Example 1 Jars of coffee

Table 1 Prices of a 100 g jar of a well-known brand of instant coffee obtained in 15 different shops in Milton Keynes on the same day in February 2012 (in pence, p)

299	315	268	269	295
295	369	275	268	295
279	268	268	295	305

There are several points to note concerning these prices.

- They relate to a particular brand of coffee. You might expect the price to vary between brands.

- They relate to a standard 100 g jar. You might expect the price per gram of this brand of coffee to vary depending upon the size of the jar – larger jars are often cheaper (per gram).
- They relate to a particular locality. You might expect the price to vary depending upon where you buy the coffee (e.g. central London, a suburb, a provincial town, a country village or a Hebridean island).
- They relate to a particular day. You might expect the price to vary from time to time depending upon changes in the cost of raw coffee beans, costs of production and distribution, and the availability of special offers.

Nevertheless, although we have data for a fixed brand of coffee, size of jar, locality and date of purchase, this batch of prices still varies from the lower extreme of 268p to the upper extreme of 369p. (In symbols: $E_L = 268$ and $E_U = 369$.) One of the most likely reasons for this is that the prices were collected from different kinds of shops (e.g. supermarket, petrol station, ethnic grocery and corner shop).

For all these reasons, it is impossible to state exactly what the price of this brand of instant coffee is. Yet its price is, in its own small way, relevant to the question: *Are people getting better or worse off?* That is, if you drink this particular coffee, then changes in its price in your locality will affect your cost of living. Similarly, your costs and economic well-being will also be affected by what happens to the prices of all the other things you need or like to consume.

On the other hand, someone who never buys instant coffee will be unaffected by any change in its price; they will be much more interested in what happens to the prices of alternative products such as ground coffee, tea, milk or fruit juice. The problem of measuring the effect of price changes on individuals with different consumption patterns will be considered in Section 5.

1.2 The median

Example 2 Picturing the coffee data

Despite the variability in the data, Table 1 does provide some idea of the price you would expect to pay for a 100 g jar of that particular instant coffee in the Milton Keynes area on that particular day. The information provided by the batch can be seen more clearly when drawn as a stemplot, and this is shown in Figure 1.

26	8 8 8 8 9
27	5 9
28	
29	5 5 5 5 9
30	5
31	5
32	
33	
34	
35	
36	9

$n = 15$ 26 | 8 represents 268 pence

Figure 1 Stemplot of coffee prices from Table 1

This shows at a glance that if you shop around, you might well find this brand of coffee on sale at less than 270p. (Indeed some stores seem to have been 'price matching' at the lowest price of 268p.) On the other hand, if you are not too careful about making price comparisons then you might pay considerably more than 300p ($3). However, you are most likely to find a shop with the coffee priced between about 270p and 300p. Although there is no one price for this coffee, it seems reasonable to say that the overall location of the price is a bit less than 300p.

The **median** of the batch is a useful measure of the overall location of the values in a batch. You met the median in the preceding unit; it was defined as the middle value of a batch of figures when the values are placed in order. Let us revise, and extend slightly, what you learned about the median in Unit 1.

See Subsection 4.2 of Unit 1.

The stemplot in Figure 1 shows the prices arranged in order of size. We can label each of these 15 prices with a symbol indicating where it comes in the ordered batch. A convenient way of showing this is to write each value as the symbol x plus a subscript number in brackets, where the subscript number shows the position of that value within the ordered batch. Figure 2 shows the 15 prices written out in ascending order using this subscript notation.

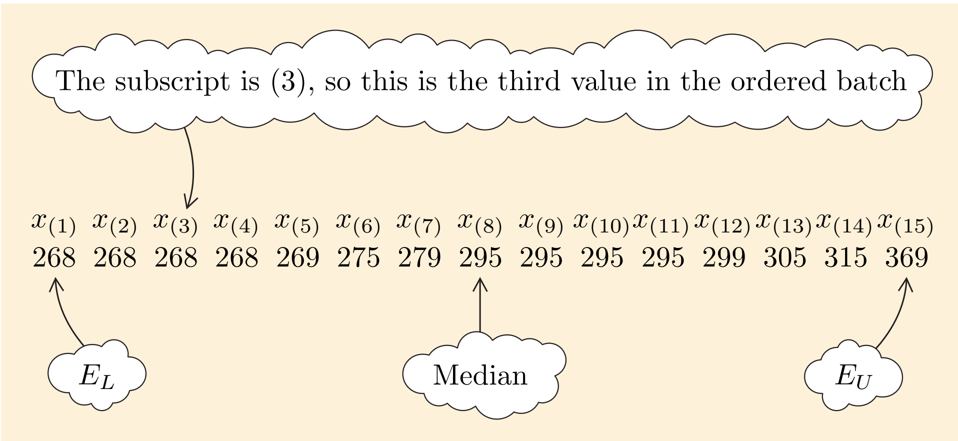


Figure 2 Subscript notation for ordered data

The lower extreme, E_L, is labelled $x_{(1)}$ and the upper extreme, E_U, is labelled $x_{(15)}$. The middle value is the value labelled $x_{(8)}$ since there are as many values, namely 7, above the value of $x_{(8)}$ as there are below it. (This is not *strictly* true here, since the values of $x_{(9)}$, $x_{(10)}$ and $x_{(11)}$ happen also to be actually equal to the median.)

This is illustrated in Figure 3 by a V-shaped formation. The median is the middle value, so it lies at the bottom of the V.

This way of picturing a batch will be developed further in Subsection 3.2.

An upside down V-shape

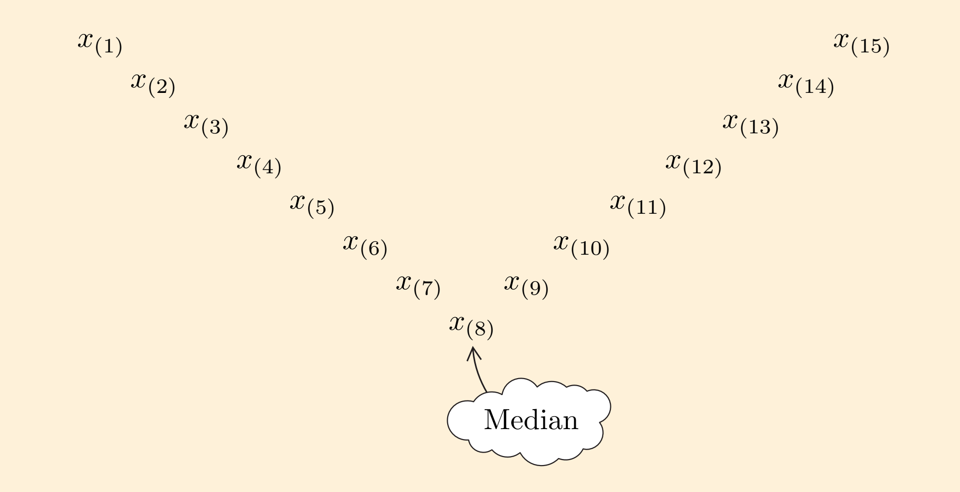


Figure 3 Median of 15 values

If you wanted to make a more explicit statement, then you could write: The median price of this batch of 15 prices is 295p.

If we picture any batch of data as a V-shape like Figure 3, the median of the batch will always lie at the bottom of the V. In the ordered batch, it is more places away from the extremes than any other value.

In general, the median is the value of the middle item when all the items of the batch are arranged in order. For a batch size n, the position of the middle value is $\frac{1}{2}(n+1)$. For example, when $n = 15$, this gives a position of $\frac{1}{2}(15+1) = 8$, indicating that $x_{(8)}$ is the median value. When n is an even number, the middle position is not a whole number and the median is the average of the two numbers either side of it. For example, when $n = 12$, the median position is $6\frac{1}{2}$, indicating that the median value is taken as halfway between $x_{(6)}$ and $x_{(7)}$.

Example 3 Digital cameras

Table 2 Prices for a particular model of digital camera as given on a price comparison website in March 2012 (to the nearest $)

60	70	53	81	74
85	90	79	65	70

If we put these prices in order and arrange them in a V-shape, they look like Figure 4.

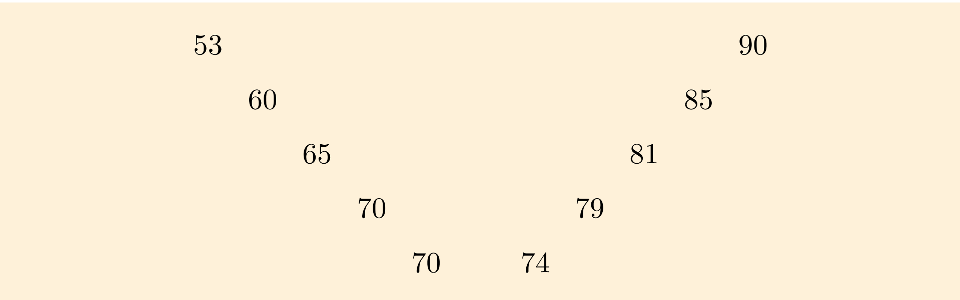


Figure 4 Prices of 10 digital cameras

Because 10 is an even number, there is no single middle value in this batch: the position of the middle item is $\frac{1}{2}(10+1) = 5\frac{1}{2}$. The two values closest to the

middle are those shown at the bottom of the V: $x_{(5)} = 70$ and $x_{(6)} = 74$. Their average is 72, so we say that the median price of this batch of camera prices is *$*72.

Activity 1 Small flat-screen televisions

Figure 5 is a stemplot of data on the prices of small flat-screen televisions. (The prices have been rounded to the nearest *$*10. Originally all but one ended in 9.99, so in this case it makes reasonable sense to ignore the rounding and treat the data as if the prices were exact multiples of *$*10.) Find the median of these data.

$$\begin{array}{r|l} 0 & 9 \\ 1 & 0 \\ 1 & 2\ 3\ 3\ 3 \\ 1 & 4\ 5\ 5\ 5\ 5 \\ 1 & 6\ 6\ 7 \\ 1 & 8\ 8\ 9 \\ 2 & \\ 2 & \\ 2 & 4\ 5 \\ 2 & 7 \end{array}$$

$n = 20 \qquad 0 \mid 9$ represents £90

Figure 5 Prices of all flat-screen televisions with a screen size of 24 inches or less on a major UK retailer's website on a day in February 2012

Not that kind of flat screen

This subsection can now be finished by using some of the methods we have met to examine a batch of data consisting of two parts, or sub-batches.

Activity 2 The price of gas in UK cities

Table 3 presents the average price of gas, in pence per kilowatt hour (kWh), in 2010, for typical consumers on credit tariffs in 14 cities in the UK. These cities have been divided into two sub-batches: as seven *northern* cities and seven *southern* cities. (Legally, at the time of writing, Ipswich is a town, not a city, but we shall ignore that distinction here.)

Table 3 Average gas prices in 14 cities

Northern		Southern	
Aberdeen	3.740	Birmingham	3.805
Edinburgh	3.740	Canterbury	3.796
Leeds	3.776	Cardiff	3.743
Liverpool	3.801	Ipswich	3.760
Manchester	3.801	London	3.818
Newcastle-upon-Tyne	3.804	Plymouth	3.784
Nottingham	3.767	Southampton	3.795

(a) Draw a stemplot of all 14 prices shown in the table.

(b) Draw separate stemplots for the seven prices for northern cities and the seven prices for southern cities.

(c) For each of these three batches (northern cities, southern cities and all cities) find the median and the range. Then use these figures to find the general level and the range of gas prices for typical consumers in the country as a whole, and to compare the north and south of the country.

Activity 2 illustrates two general properties of sub-batches:

- The *range* of the complete batch is greater than or equal to the ranges of all the sub-batches.
- The *median* of the complete batch is greater than or equal to the smallest median of a sub-batch and less than or equal to the largest median of a sub-batch.

1.3 The arithmetic mean

Another important measure of location is the arithmetic mean. (Pronounced arith*met*ic.)

Arithmetic mean

The arithmetic mean is the sum of all the values in the batch divided by the size of the batch. More briefly,

$$\text{mean} = \frac{\text{sum}}{\text{size}}.$$

There are other kinds of mean, such as the geometric mean and the harmonic mean, but in this module we shall be using only the arithmetic mean; the word *mean* will therefore normally be used for *arithmetic mean*.

Example 4 An arithmetic mean

Suppose we have a batch consisting of five values: 4, 8, 4, 2, 9. In this simple example, the mean is

$$\frac{\text{sum}}{\text{size}} = \frac{4+8+4+2+9}{5} = \frac{27}{5} = 5.4.$$

Note that in calculating the mean, the order in which the values are summed is irrelevant.

For a larger batch size, you may find it helpful to set out your calculations systematically in a table. However, in practice the raw data are usually fed directly into a computer or calculator. In general, it is a good idea to check your calculations by reworking them. If possible, use a different method in the reworking; for example, you could sum the numbers in the opposite order.

The formula 'mean $=$ sum/size' can be expressed more concisely as follows. Referring to the values in the batch by x, the 'sum' can be written as $\sum x$. Here $\sum$ is the Greek (capital) letter Sigma, the Greek version of S, and is used in statistics to denote 'the sum of'. Also, the symbol $\overline{x}$ is often used to denote the mean – and as you have already seen in stemplots, n can be used to denote the

batch size. (Some calculators use keys marked $\sum x$ and $\overline{x}$ to produce the sum and the mean of a batch directly.)

Using this notation,

$$\text{mean} = \frac{\text{sum}}{\text{size}}$$

can be written as

$$\overline{x} = \frac{\sum x}{n}.$$

In this module we shall normally round the mean to one more figure than the original data.

Activity 3 Small televisions: the mean

The prices of 20 small televisions were given in Activity 1 (Subsection 1.2). Find the mean of these prices. Round your answer appropriately (if necessary), given that the original data were rounded to the nearest \$10.

1.4 The mean and median compared

Both the mean and median of a batch are useful indicators of the location of the values in the batch. They are, however, calculated in very different ways. To find the median you must first order the batch of data, and if you are not using a computer, you will often do the sorting by means of a stemplot. On the other hand, the major step in finding the mean consists of summing the values in the batch, and for this they do not need to be ordered.

For large batches, at least when you are not using a computer, it is often much quicker to sum the values in the batch than it is to order them. However, for small batches, like some of those you will be analysing in this module without a computer, it can be just as fast to calculate the median as it is to calculate the mean. Moreover, placing the batch values in order is not done solely to help calculate the median – there are many other uses. Drawing a stemplot to order the values also enables us to examine the general shape of the batch, as you saw in Unit 1. In Section 3 you will read about some other uses of the stemplot.

Comparisons based on the method of calculation can be of great practical interest, but the rest of this subsection will consider more fundamental differences between the mean and the median – differences which should influence you when you are deciding which measure to use in summarising the general location of the values in a batch.

Many of the problems with the mean, as well as some advantages, lie in the fact that the precise value of *every* item in the batch enters into its calculation. In calculating the median, most of the data values come into the calculation only in terms of whether they are in the 50% above the median value or the 50% below it. If one of them changes slightly, but without moving into the other half of the batch, the median will not change. In particular, if the extreme values in the batch are made smaller or larger, this will have no effect on the value of the median – the median is resistant to outliers, as noted in Unit 1. In contrast, changes to the extremes could have an appreciable effect on the value of the mean, as the following examples show.

Example 5 Changing the extreme coffee prices

For the batch of coffee prices in Figure 1 (Subsection 1.2), the sum of the values is 4363p, so the mean is

$$\frac{4363\text{p}}{15} \simeq 290.9\text{p}.$$

Suppose the highest and lowest coffee prices are reduced so that

$$x_{(1)} = 240 \quad \text{and} \quad x_{(15)} = 340.$$

The median of this altered batch is the same as before, 295p. However, the sum of the values is now 4306p and so the mean is

$$\frac{4306\text{p}}{15} \simeq 287.1\text{p}.$$

Example 6 Changing the small television prices

Suppose the highest two television prices in Activity 1 (Subsection 1.2) are altered to \$350 and \$400. The median, at \$150, remains the same as that of the original batch, whereas the new mean is

$$\frac{\$3470}{20} = \$173.5 \simeq \$174$$

compared with the original mean of \$162.

Now, even with the very high prices of \$350 and \$400 for two televisions, the overall location of the main body of the data is still much the same as for the original batch of data. For the original batch the mean, \$162, was a reasonably good measure of this. However, for the new batch the mean, \$174, is much too high to be a representative measure since, as we can see from the stemplot in Activity 1, most of the values are below \$174.

Example 6 is the subject of Screencast 1 for Unit 2 (see the module website).

The idea of resistance to outliers was introduced in Subsection 4.2 of Unit 1.

A measure which is insensitive to changes in the values near the extremes is called a **resistant measure**.

The *median* is a **resistant** measure whereas the *mean* is **sensitive**.

In the following activities, you can investigate some other ways in which the median is *more* resistant than the mean.

Activity 4 Changing the gas prices

In Activity 2 (Subsection 1.2) you may have noticed that Cardiff and Ipswich had rather low gas prices compared to the other southern cities. Here you are going to examine the effect of deleting them from the batch of southern cities. Complete the following table and comment on your results.

Batch	Mean	Median
Seven southern cities		
Five southern cities (excluding Cardiff and Ipswich)		

Activity 5 A misprint in the gas prices

Suppose the value for London had been misprinted as 8.318 instead of 3.818 (quite an easy mistake to make!). How would this affect your results for the batch of five southern cities (again omitting Cardiff and Ipswich)?

Batch	Mean	Median
Five cities (correct data)		
Five cities (with misprint)		

Suppose you wanted to use these values – the correct ones, of course – to estimate the average price of gas over the whole country. The simple arithmetic mean of the 14 values given in Table 3 would not allow for the fact that much more gas is consumed in London, at a relatively high price, than in other cities. To take account of this you would need to calculate what is known as a *weighted* arithmetic mean. Weighted means are the subject of the next section.

Exercises on Section 1

Exercise 1 Finding medians

For each of the following batches of data, find the median of the batch. (We shall also use these batches of data in some of the exercises in Section 3; they come from Figure 37 and Table 11 of Unit 1 (towards the end of Subsections 5.2 and 5.1 respectively).)

(a) Percentage scores in arithmetic:

0	7
1	5
2	
3	3 5
4	2 2 3
5	5 8
6	4 6 8
7	1 1 6 8 9
8	0 1 1 3 4 5 5 6 9
9	1 1 3 5 9
10	0 0

$n = 33$ 0 | 7 represents a score of 7%

(b) Prices of 26 digital televisions ($):

170	180	190	200	220	229	230	230	230
230	250	269	269	270	279	299	300	300
315	320	349	350	400	429	649	699	

Exercise 2 Finding means

Calculate the mean for each of the batches in Exercise 1.

Exercise 3 The effect of removing values on the median and mean

In the data on prices for small televisions in Activity 1 (Subsection 1.2), the three highest-priced televisions were considerably more expensive than all the others (which all cost under $200). Suppose that in fact these prices had been for a different, larger type of television that should not have been in the batch. (In fact that is not the case – but this is only an exercise!) Leave these three prices out of the batch and calculate the median and the mean of the remaining prices.

How do these values compare with the original median (150) and mean (162)? What does this comparison demonstrate about how resistant the median and mean are?

You have now covered the material needed for Subsection 2.1 of the Computer Book.

2 Weighted means

For goods and services, price changes vary considerably from one to another. Central to the theme question of this unit and the next, *Are people getting better or worse off?*, there is a need to find a fair method of calculating the average price change over a wide range of goods and services. Clearly a 10% rise in the price of bread is of greater significance to most people than a similar rise in the price of clothes pegs, say. What we need to take account of, then, are the relative *weightings* attached to the various price changes under consideration.

2.1 The mean of a combined batch

This first subsection looks at how a mean can be calculated when two unequally weighted batches are combined.

Example 7 Alan's and Beena's biscuits

Suppose we are conducting a survey to investigate the general level of prices in some locality. Two colleagues, Alan and Beena, have each visited several shops and collected information on the price of a standard packet of a particular brand of biscuits. They report as follows (Figure 6).

- Alan visited five shops, and calculated that the mean price of the standard packet at these shops was 81.6p.
- Beena visited eight shops, and calculated that the mean price of the standard packet at these shops was 74.0p.

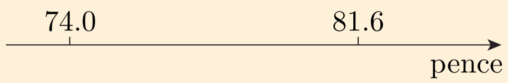


Figure 6 Means of biscuit prices

If we had all the individual prices, five from Alan and eight from Beena, then they could be amalgamated into a single batch of 13 prices, and from this combined batch we could calculate the mean price of the standard packet at all 13 shops.

However, our two investigators have unfortunately not written down, nor can they fully remember, the prices from individual shops. Is there anything we can do to calculate the mean of the combined batch?

Fortunately there is, as long as we are interested in arithmetic means. (If they had recorded the medians instead, then there would have been very little we could do.)

The mean of the combined batch of all 13 prices will be calculated as

$$\frac{\text{sum (of the combined batch prices)}}{\text{size (of the combined batch)}}.$$

We already know that the size of the combined batch is the sum of the sizes of the two original batches; that is, $5 + 8 = 13$. The problem here is how to find the sum of the combined batch of Alan's and Beena's prices. The solution is to rearrange the familiar formula

$$\text{mean} = \frac{\text{sum}}{\text{size}}$$

so that it reads

$$\text{sum} = \text{mean} \times \text{size}.$$

This will allow us to find the sums of Alan's five prices and Beena's eight prices separately. Adding the results will produce the sum of the combined batch prices. Finally, dividing by 13 completes the calculation of finding the combined batch mean.

Let us call the sum of Alan's prices 'sum(A)' and the sum of Beena's prices 'sum(B)'.

For Alan: mean $= 81.6$ and size $= 5$, so sum(A) $= 81.6 \times 5 = 408$.

For Beena: mean $= 74.0$ and size $= 8$, so sum(B) $= 74.0 \times 8 = 592$.

For the combined batch:

$$\begin{aligned}\text{mean} &= \frac{\text{combined sum}}{\text{combined size}}\\ &= \frac{408 + 592}{13}\\ &= \frac{1000}{13} \simeq 76.9\end{aligned}$$

Here, the result has been rounded to give the same number of digits as in the two original means.

The process that we have used above is an important one. It will be used several times in the rest of this unit. The box below summarises the method, using symbols.

Mean of a combined batch

The formula for the **mean** $\overline{x}_C$ **of a combined batch** C is

$$\overline{x}_C = \frac{\overline{x}_A n_A + \overline{x}_B n_B}{n_A + n_B},$$

where batch C consists of batch A combined with batch B, and

$\overline{x}_A =$ mean of batch A, $\quad n_A =$ size of batch A,

$\overline{x}_B =$ mean of batch B, $\quad n_B =$ size of batch B.

For our survey in Example 7,

$$\overline{x}_A = 81.6, \quad n_A = 5, \quad \overline{x}_B = 74.0, \quad n_B = 8.$$

The formula summarises the calculations we did as

$$\overline{x}_C = \frac{(81.6 \times 5) + (74.0 \times 8)}{5 + 8}.$$

This expression is an example of a **weighted mean**. The numbers 5 and 8 are the **weights**. We call this expression the weighted mean of 81.6 and 74.0 with weights 5 and 8, respectively.

To see why the term *weighted mean* is used for such an expression, imagine that Figure 7 shows a horizontal bar with two weights, of sizes 5 and 8, hanging on it at the points 81.6 and 74.0, and that you need to find the point at which the bar will balance. This point is at the weighted mean: approximately 76.9.

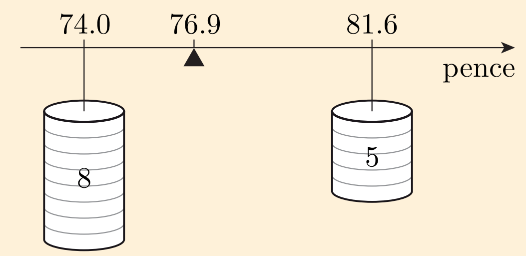


Figure 7 Point of balance at the weighted mean

This physical analogy illustrates several important facts about weighted means.

- It does not matter whether the weights are 5 kg and 8 kg or 5 tonnes and 8 tonnes; the point of balance will be in the same place. It will also remain in the same place if we use weights of 10 kg and 16 kg or 40 kg and 64 kg – it is only the *relative sizes* (i.e. the *ratio*) of the weights that matter.
- The point of balance must be between the points where we hang the weights, and it is nearer to the point with the larger weight.
- If the weights are equal, then the point of balance is halfway between the points.

This gives the following rules.

Rules for weighted means

Rule 1 The weighted mean depends on the relative sizes (i.e. the ratio) of the weights.

Rule 2 The weighted mean of two numbers always lies between the numbers and it is nearer the number that has the larger weight.

Rule 3 If the weights are equal, then the weighted mean of two numbers is the number halfway between them.

Example 8 Two batches of small televisions

Suppose that we have two batches of prices (in pounds) for small televisions:

Batch A has mean 119 and size 7.

Batch B has mean 185 and size 13.

To find the mean of the combined batch we use the formula above, with

$$\overline{x}_A = 119, \quad n_A = 7, \quad \overline{x}_B = 185, \quad n_B = 13.$$

This gives

$$\begin{aligned}\overline{x}_C &= \frac{(119 \times 7) + (185 \times 13)}{7 + 13} \\ &= \frac{833 + 2405}{20} \\ &= \frac{3238}{20} \\ &= 161.9 \simeq 162.\end{aligned}$$

Note that this is the weighted mean of 119 and 185 with weights 7 and 13 respectively. It lies between 119 and 185 but it is nearer to 185 because this has the greater weight: 13 compared with 7.

Example 8 is the subject of Screencast 2 for Unit 2 (see the module website).

You have also now covered the material needed for Subsection 2.2 of the Computer Book.

2.2 Further uses of weighted means

We shall now look at another similar problem about mean prices – one which is perhaps closer to your everyday experience.

Example 9 Buying petrol

Suppose that, in a particular week in 2012, a motorist purchased petrol on two occasions. On the first she went to her usual, relatively low-priced filling station where the price of unleaded petrol was 136.9p per litre and she filled the tank; the quantity she purchased was 41.2 litres. The second occasion saw her obliged to purchase petrol at an expensive service station where the price of unleaded petrol was 148.0p per litre; she therefore purchased only 10 litres. What was the mean price, in pence per litre, of the petrol she purchased during that week?

To calculate this mean price we need to work out the total expenditure on petrol, in pence, and divide it by the total quantity of petrol purchased, in litres.

The total quantity purchased is straightforward as it is just the sum of the two quantities, so $41.2 + 10$.

To find the expenditure on each occasion, we need to apply the formula:

$$\text{cost} = \text{price} \times \text{quantity}.$$

This gives 136.9×41.2 and 148.0×10, respectively.

So the total expenditure, in pence, is $(136.9 \times 41.2) + (148.0 \times 10)$. The mean price, in pence per litre, for which we were asked, is this total expenditure divided by the total number of litres bought:

$$\frac{(136.9 \times 41.2) + (148.0 \times 10)}{41.2 + 10}.$$

We have left the answer in this form, rather than working out the individual products and sums as we went along, to show that it has the same form as the calculation of the combined batch mean. (The answer is 139.07p per litre, rounded from 139.067 97p per litre.)

The phrase 'goods and services' is an awkward way of referring to the things that are relevant to the cost of living; that is, physical things you might buy, such as bread or gas, and services that you might pay someone else to do for you, such as window-cleaning. Economists sometimes use the word *commodity* to cover both goods and services that people pay for, and we shall use that word from time to time in this unit. (Note that there are other, different, technical meanings of commodity that you might meet in different contexts.)

The mean price of a quantity bought on two different occasions

In general, if you purchase q_1 units of some commodity at p_1 pence per unit and q_2 units of the same commodity at p_2 pence per unit, then the mean price of this commodity, $\overline{p}$ pence per unit, can be calculated from the following formula:

$$\overline{p} = \frac{p_1\,q_1 + p_2\,q_2}{q_1 + q_2}.$$

Example 10 Buying potatoes

Suppose that, in one month, a family purchased potatoes on two occasions. On one occasion they bought 10 kg at 40p per kg, and on another they bought 6 kg at 45p per kg. We can use this formula to calculate the mean price (in pence per kg) that they paid for potatoes in that month. We have

$$\left.\begin{array}{ll} q_1 = 10 & \text{quantity} \\ p_1 = 40 & \text{price} \end{array}\right\} \text{first occasion}$$

and

$$\left.\begin{array}{ll} q_2 = 6 & \text{quantity} \\ p_2 = 45 & \text{price} \end{array}\right\} \text{second occasion.}$$

This gives

$$\begin{aligned} \overline{p} &= \frac{(40 \times 10) + (45 \times 6)}{10 + 6} \\ &= \frac{400 + 270}{16} \\ &= \frac{670}{16} \\ &= 41.875 \simeq 41.9. \end{aligned}$$

So the mean price for that month is 41.9p per kg.

The two formulas we have been using,

$$\frac{\overline{x}_A n_A + \overline{x}_B n_B}{n_A + n_B} \quad \text{and} \quad \frac{p_1 q_1 + p_2 q_2}{q_1 + q_2},$$

are basically the same; they are both examples of weighted means.

The first formula is the weighted mean of the numbers $\overline{x}_A$ and $\overline{x}_B$, using the batch sizes, n_A and n_B, as weights.

The second formula is the weighted mean of the unit prices p_1 and p_2, using the quantities bought, q_1 and q_2, as weights.

The general form of a weighted mean of two numbers having associated weights is as follows.

Weighted mean of two numbers

The **weighted mean** of the two numbers x_1 and x_2 with corresponding weights w_1 and w_2 is

$$\frac{x_1 w_1 + x_2 w_2}{w_1 + w_2}.$$

Weighted means have many uses, two of which you have already met. The type of weights depends on the particular use. In our uses, the weights were the following.

- The sizes of the batches, when we were calculating the combined batch mean from two batch means.
- The quantities bought, when we were calculating the mean price of a commodity bought on two separate occasions.

Another very important use is in the construction of an index, such as the Retail Prices Index; we shall therefore be making much use of weighted means in the final sections of this unit.

In the next example, we do not have all the information required to calculate the mean, but we can still get a reasonable answer by using weights.

Example 11 Weighted means of two gas prices

Let us return to the gas prices in Table 3 (Subsection 1.2). This has information about the price of gas for typical consumers in individual cities, but no national figure. Suppose that you want to combine these figures to get an average figure for the whole country; how could you do it? At the end of Section 1, it was suggested that weighted means could provide a solution. The complete answer to this question, using weighted means, is in Example 13 towards the end of this section. To introduce the method used there, let us now consider a similar, but simpler, question.

Here we use just two cities, London and Edinburgh, where the prices were 3.818p per kWh and 3.740p per kWh respectively. How can we combine these two values into one sensible average figure?

One possibility would be to take the simple mean of the two numbers. This gives

$$\tfrac{1}{2}(3.818 + 3.740) = 3.779.$$

However, this gives both cities *equal* weight. Because London is a lot larger than Edinburgh, we should expect the average to be nearer the London price than the Edinburgh price.

This suggests that we use a *weighted* mean of the form

$$\frac{3.818q_1 + 3.740q_2}{q_1 + q_2},$$

where q_1 and q_2 are suitably chosen weights, with the weight q_1 of the London price larger than the weight q_2 of the Edinburgh price.

The best weights would be the total quantities of gas consumed in 2010 in each city. However, even if this information is not available to us, we can still find a reasonable average figure by using as weights a readily available measure of the sizes of the two cities: their populations.

The populations of the urban areas of these cities are approximately 8 300 000 and 400 000 respectively. So we could put $q_1 = 8\,300\,000$ and $q_2 = 400\,000$.

However, we know that the weighted mean depends only on the ratio of the weights. Therefore, the weights $q_1 = 83$ and $q_2 = 4$ will give the same answer.

These weights give

$$\frac{(3.818 \times 83) + (3.740 \times 4)}{83 + 4}.$$

Activity 6 Using the rules for weighted means

Using the rules for weighted means, would you expect the weighted mean price to be nearer the London price or the Edinburgh price? To check, calculate the weighted mean price.

Although we cannot think of the weighted mean price in Activity 6 as a calculation of the total cost divided by the total consumption, the answer *is* an estimate of the average price, in pence per kWh, for typical consumers in the two cities, and it is the best estimate we can calculate with the available information.

Sometimes the weights in a weighted mean do not have any significance in themselves: they are neither quantities, nor sizes, etc., but simply weights. This is illustrated in the following activity.

Activity 7 Weighted means of Open University marks

As an Open University student, an example of the use of weighted means with which you are familiar, or will soon become familiar, is the combination of interactive computer-marked assignment (iCMA) and tutor-marked assignment (TMA) scores to provide an overall continuous assessment score (OCAS).

Suppose that you obtain a score of 80 for your iCMAs and a score of 60 for your TMAs. (I am not saying these are typical scores for M140!) Calculate what your overall continuous assessment score will be if the weights for the two components are as follows.

(a) iCMA 50, TMA 50

(b) iCMA 40, TMA 60

(c) iCMA 65, TMA 55

(d) iCMA 25, TMA 75

(e) iCMA 30, TMA 90

This is Rule 1 for weighted means (see Subsection 2.1).

We have seen, in Activity 7 and in Example 11, that only the ratio of the weights affects the answer, not the individual weights. So weights are often chosen to add up to a convenient number like 100 or 1000.

Activity 7 should also have reminded you of another important property of a weighted mean of two numbers: the weighted mean lies nearer to the number having the larger weight.

This is part means.

2.3 More than two numbers

The idea of a weighted mean can be extended to more than two numbers. To see how the calculation is done in general, remind yourself first how we calculated the weighted mean of two numbers x_1 and x_2 with corresponding weights w_1 and w_2.

1. Multiply each number by its weight to get the products x_1w_1 and x_2w_2.
2. Sum these products to get $x_1w_1 + x_2w_2$.
3. Sum the weights to get $w_1 + w_2$.
4. Divide the sum of the products by the sum of the weights.

This leads to the following formula.

Weighted mean of two or more numbers

The weighted mean of two or more numbers is

$$\frac{\text{sum of }\{\text{number} \times \text{weight}\}}{\text{sum of weights}} = \frac{\text{sum of products}}{\text{sum of weights}}.$$

This is the formula which is used to find the weighted mean of any set of numbers, each with a corresponding weight.

Example 12 A weighted mean of wine prices

Suppose we have the following three batches of wine prices (in pence per bottle).

Batch 1 with mean 525.5 and batch size 6.

Batch 2 with mean 468.0 and batch size 2.

Batch 3 with mean 504.2 and batch size 12.

We want to calculate the weighted mean of these three batch means using, as corresponding weights, the three batch sizes. Rather than applying the formula directly, the calculations can be set out in columns.

Table 4 Data on wine purchases

Batch	Number (batch mean)	Weight (batch size)	Number × weight (= product)
Batch 1	525.5	6	3 153.0
Batch 2	468.0	2	936.0
Batch 3	504.2	12	6 050.4
Sum		20	10 139.4

The weighted mean is

$$\frac{\text{sum of products}}{\text{sum of weights}} = \frac{10\,139.4}{20} = 506.97.$$

We round this to the same accuracy as the original means, to get a weighted mean of 507.0. (Note that this lies between 468.0 and 525.5. This is a useful check, as a weighted mean always lies within the range of the original means.)

The physical analogy in Example 12 can be extended to any set of numbers and weights. Suppose that you calculate the weighted mean for:

1.3 with weight 2

1.9 with weight 1

1.7 with weight 3.

This is given by

$$\frac{(1.3 \times 2) + (1.9 \times 1) + (1.7 \times 3)}{2 + 1 + 3} = \frac{2.6 + 1.9 + 5.1}{6} = \frac{9.6}{6} = 1.6.$$

This is pictured in Figure 8, with the point of balance for these three weights shown at 1.6.

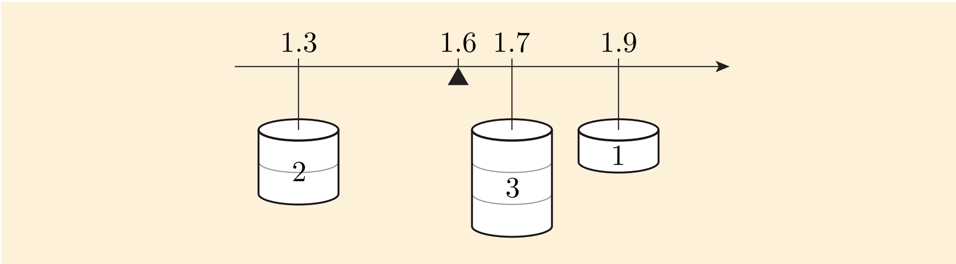


Figure 8 Point of balance for three means

You will meet many examples of weighted means of larger sets of numbers in Subsection 5.2, but we shall end this section with one more example.

Example 13 Weighted means of many gas prices

Example 11 showed the calculation of a weighted mean of gas prices using, for simplicity, just the two cities London and Edinburgh. We can extend Example 11 to calculate a weighted mean of all 14 gas prices from Table 3, using as weights the populations of the 14 cities. The calculations are set out in Table 5.

Table 5 Product of gas price and weight by city

	Price (p/kWh) x	Weight w	Price $\times$ weight xw
Aberdeen	3.740	19	71.060
Edinburgh	3.740	42	157.080
Leeds	3.776	150	566.400
Liverpool	3.801	82	311.682
Manchester	3.801	224	851.424
Newcastle-upon-Tyne	3.804	88	334.752
Nottingham	3.767	67	252.389
Birmingham	3.805	228	867.540
Canterbury	3.796	5	18.980
Cardiff	3.743	33	123.519
Ipswich	3.760	14	52.640
London	3.818	828	3161.304
Plymouth	3.784	24	90.816
Southampton	3.795	30	113.850
Sum		1834	6973.436

The entries in the weight column, w, are the approximate populations, in 10 000s, of the urban areas that include each city (as measured in the 2001 Census). For each city, we multiply the price, x, by the weight, w, to get the entry in the last column, xw.

The weighted mean of the gas prices using these weights is then

$$\frac{\text{sum of products (price} \times \text{weight)}}{\text{sum of weights}}$$

or, in symbols,

$$\frac{\sum xw}{\sum w}.$$

As $\sum xw = 6973.436$ and $\sum w = 1834$, the weighted mean is

$$\frac{6973.436}{1834} = 3.802\,310 \simeq 3.802.$$

So the weighted mean of these gas prices, using approximate population figures as weights, is 3.802p per kWh.

Note that this weighted mean is larger than all but three of the gas prices for individual cities. That is because the cities with the two highest populations, London and Birmingham, also have the highest gas prices, and the weighted mean gas price is pulled towards these high prices.

Although the details of the calculation above are written out in full in Table 5, in practice, using even a simple calculator, this is not necessary. It is usually possible to keep a running sum of both the weights and the products as the data are being entered. One way of doing this is to accumulate the sum of the weights into the calculator's memory while the sum of the products is cumulated on the display. If you are using a specialist statistics calculator, the task is generally very straightforward. Simply enter each price and its corresponding weight using the method described in your calculator instructions for finding a weighted mean.

Activity 8 Weighted means on your calculator

Use your calculator to check that the sum of weights and sum of products of the data in Table 5 are, respectively, 1834 and 6973.436, and that the weighted mean is 3.802. (No solution is given to this activity.)

Activity 9 Weighted mean electricity price

Table 6 is similar to Table 5, but this time it presents the average price of *electricity*, in pence per kilowatt hour (kWh). These data are again for the year 2010 for typical consumers on credit tariffs in the same 14 cities we have been considering for gas prices, with the addition of Belfast. Again, the weights are the approximate populations of the relevant urban areas, in 10 000s.

Table 6 Populations and electricity prices in 15 cities

	Price (p/kWh) x	Weight w	Price $\times$ weight xw
Aberdeen	13.76	19	
Belfast	15.03	58	
Edinburgh	13.86	42	
Leeds	12.70	150	
Liverpool	13.89	82	
Manchester	12.65	224	
Newcastle-upon-Tyne	12.97	88	
Nottingham	12.64	67	
Birmingham	12.89	228	
Canterbury	12.92	5	
Cardiff	13.83	33	
Ipswich	12.84	14	
London	13.17	828	
Plymouth	13.61	24	
Southampton	13.41	30	
Sum			

Use these data to calculate the weighted mean electricity price. (Your calculator will almost certainly allow you to do this without writing out all the values in the xw column.)

Exercises on Section 2

Exercise 4 A combined batch of camera prices

Find the mean price of the batch formed by combining the following two batches, A and B, of camera prices.

Batch A has mean price $\$80.7$ and batch size 10.

Batch B has mean price $\$78.5$ and batch size 17.

Exercise 5 The mean price of fabric

Suppose you buy 8.5 metres of fabric in a sale, at \$10.95 per metre, to make some bedroom curtains. The following year you decide to make a matching bedspread and so you buy 6 metres of the same material, but the price is now \$12.70 per metre. Calculate the mean price of all the material, in \$ per metre.

I like to sleep each night with my feet in the oven and my head in the freezer. That way I'm comfortable on average.

3 Measuring spread

As you have already seen, it is difficult to measure price changes when they so often vary from shop to shop and region to region. Taking some average value, such as the median or the mean, helps to simplify the problem. However, it would be a mistake to ignore the notion of spread, as averages on their own can be misleading.

Information about spread can be very important in statistical analysis, where you are often interested in comparing two or more batches. In this section we shall

look first at measures of spread, and then at some methods of summarising the shape of a batch of data.

But how can spread be measured? Just as there are several ways of measuring location (mean, median, etc.), there are also several ways of measuring spread. Here, we shall examine two such measures: the *range* and the *interquartile range*.

In the next unit you will learn about a further measure of spread called the *standard deviation*.

3.1 The range

You have already met the range, which is defined below.

The range

The range is the distance between the lower and the upper extremes. It can be calculated from the formula:

$$\text{range} = E_U - E_L,$$

where E_U is the upper extreme and E_L is the lower extreme.

See Subsection 4.2 of Unit 1.

Given an ordered batch of data, for example in a stemplot, the range can easily be calculated. However, the range tells us very little about how the values in the main body of the data are spread. It is also very sensitive to changes in the extreme values, like those considered in Subsection 1.4. It would be better to have a measure of spread that conveys more information about the spread of values in the main body of the data. One such measure is based upon the difference between two particular values in the batch, known as the **quartiles**. As the name suggests, the two quartiles lie one quarter of the way into the batch from either end. The major part of the next subsection describes how to find them.

More birds, now showing the shape of the /\/\ diagram

3.2 Quartiles and the interquartile range

Finding the quartiles of a batch is very similar to finding the median.

In Subsection 1.2, we represented a batch as a V-shaped formation, with the median at the 'hinge' where the two arms of the V meet. The median splits the batch into two equal parts. Similarly, we can put another hinge in each side of the V and get four roughly equal parts, shaped like this: /\/\. For a batch of size 15, it looks like Figure 9.

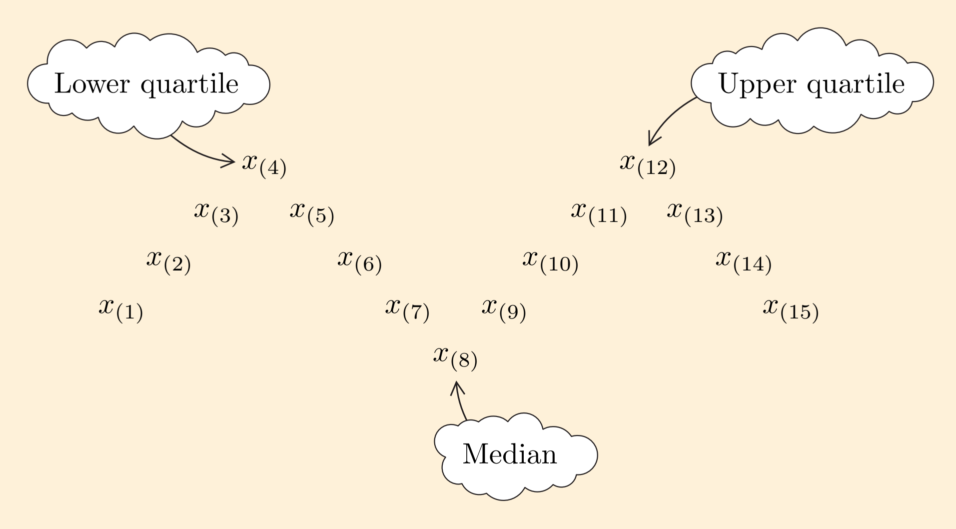


Figure 9 Median and quartiles

The points at the side hinges, in this case $x_{(4)}$ and $x_{(12)}$, are the quartiles. There are two quartiles which, as with the extremes, we call the **lower quartile** and the **upper quartile**. The lower quartile separates off the bottom quarter, or lowest 25%. The upper quartile separates off the top quarter, or highest 25%. They are denoted Q_1 and Q_3 respectively. (Sometimes they are referred to as the *first quartile* and the *third quartile*.)

You might be wondering, if these are Q_1 and Q_3, what happened to Q_2? Well, have a think about that for a moment.

Q_1 separates the bottom quarter of the data (from the top three quarters), and Q_3 separates the bottom three quarters (from the top quarter). So it would make sense to say that Q_2 separates the bottom two quarters (from the top two quarters). But two quarters make a half, so Q_2 would denote the median, and since there is already a separate word for that, it's not usual to call it the second quartile.

Usually we cannot divide the batch exactly into quarters. Indeed, this is illustrated in Figure 9 where the two central parts of the $\wedge\!\wedge$ are larger than the outer ones. As with calculating the median for an even-sized batch, some rule is needed to tell us how many places we need to count along from the smallest value to find the quartiles. However, there are several alternatives that we could adopt and the particular rule described below is somewhat arbitrary. Different authors and different software may use slightly different rules. The rule adopted here is the one used by Minitab. If your calculator can find quartiles, note that it may use a different rule, and you may also have used different rules in other Open University modules.

As you might have expected, the rule involves dividing $(n + 1)$ by 4, where n is the batch size (as opposed to dividing by 2 to find the median). However, the rule is slightly more complicated for the quartiles and it depends on whether n is exactly divisible by 4.

The quartiles

The lower quartile Q_1 is at position $\dfrac{(n+1)}{4}$ in the ordered batch.

The upper quartile Q_3 is at position $\dfrac{3(n+1)}{4}$ in the ordered batch.

If $(n+1)$ is exactly divisible by 4, these positions correspond to a single value in the batch.

If $(n+1)$ is *not* exactly divisible by 4, then the positions are to be interpreted as follows.

- A position which is a whole number followed by $\frac{1}{2}$ means 'halfway between the two positions either side' (as was the case for finding the median).
- A position which is a whole number followed by $\frac{1}{4}$ means 'one quarter of the way from the position below to the position above'. So for instance if a position is $5\frac{1}{4}$, the quartile is the number one quarter of the way from $x_{(5)}$ to $x_{(6)}$.
- A position which is a whole number followed by $\frac{3}{4}$ means 'three quarters of the way from the position below to the position above'. So for instance if a position is $4\frac{3}{4}$, the quartile is the number three quarters of the way from $x_{(4)}$ to $x_{(5)}$.

Before we actually use these rules to find quartiles, let us look at some more examples of $\wedge\wedge$-shaped diagrams for different batch sizes n. The case where $(n+1)$ is exactly divisible by 4, so that $\frac{1}{4}(n+1)$ is a whole number, was shown in Figure 9. The following three figures show the three other possible scenarios, where $(n+1)$ is not exactly divisible by 4.

For $n = 17$, $\frac{1}{4}(n+1) = 4\frac{1}{2}$ and $\frac{3}{4}(n+1) = 13\frac{1}{2}$. So Q_1 is halfway between $x_{(4)}$ and $x_{(5)}$, and Q_3 is halfway between $x_{(13)}$ and $x_{(14)}$.

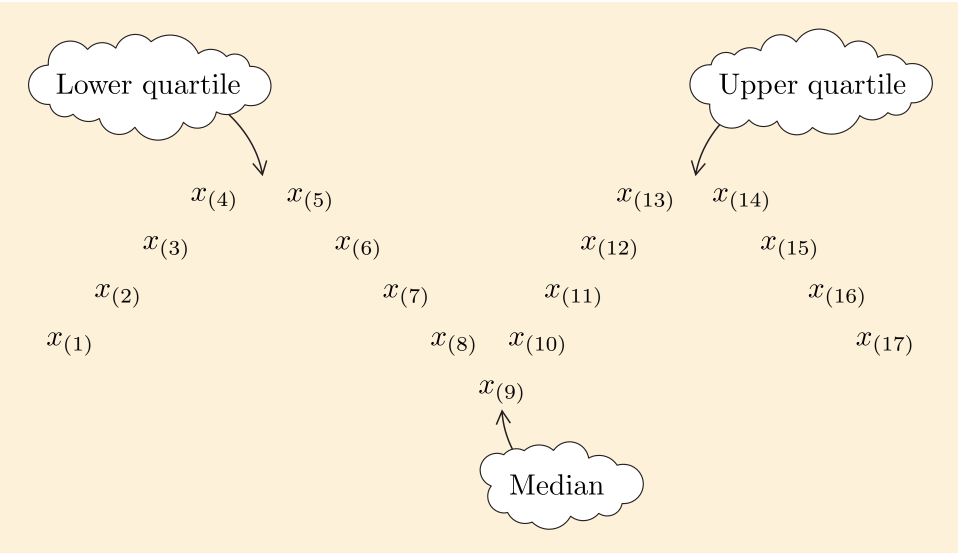


Figure 10 Quartiles for sample size $n = 17$

For $n = 18$, $\frac{1}{4}(n+1) = 4\frac{3}{4}$ and $\frac{3}{4}(n+1) = 14\frac{1}{4}$. So Q_1 is three quarters of the way from $x_{(4)}$ to $x_{(5)}$, and Q_3 is one quarter of the way from $x_{(14)}$ to $x_{(15)}$.

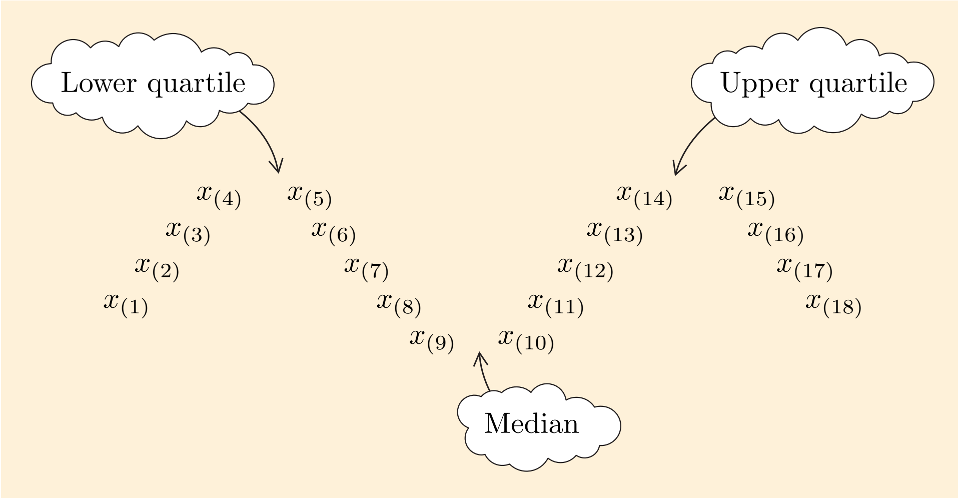


Figure 11 Quartiles for sample size $n = 18$

For $n = 20$, $\frac{1}{4}(n+1) = 5\frac{1}{4}$ and $\frac{3}{4}(n+1) = 15\frac{3}{4}$. So Q_1 is one quarter of the way from $x_{(5)}$ to $x_{(6)}$, and Q_3 is three quarters of the way from $x_{(15)}$ to $x_{(16)}$.

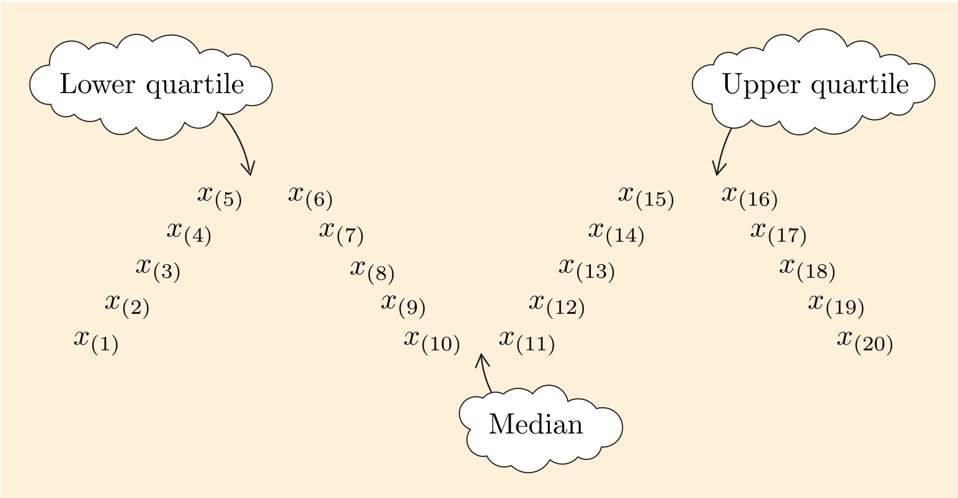


Figure 12 Quartiles for sample size $n = 20$

Example 14 Quartiles for the prices of small televisions

Figure 12 showed you where the quartiles are for a batch of size 20. Let us now use the stemplot of the 20 television prices in Figure 13, which you first met in Figure 5 (Subsection 1.2), to find the lower and upper quartiles, Q_1 and Q_3, of this batch.

0	9
1	0
1	2 3 3 3
1	4 5 5 5 5
1	6 6 7
1	8 8 9
2	
2	
2	4 5
2	7

$n = 20$ 0 | 9 represents £90

Figure 13 Prices of flat-screen televisions with a screen size of 24 inches or less

To calculate the lower quartile Q_1 you need to find the number that is one quarter of the way from $x_{(5)}$ to $x_{(6)}$. These values are both 130, so Q_1 is 130. To calculate the upper quartile Q_3 you need to find the number three quarters of the way from $x_{(15)}$ to $x_{(16)}$. These values are both 180, so Q_3 is 180.

That example was easier than it might have been, because for each quartile the two numbers we had to consider turned out to be equal!

Example 15 Quartiles for the camera prices

Table 2 (Subsection 1.2) gave ten prices for a particular model of digital camera (in pounds). In order, the prices are as follows.

$$53 \quad 60 \quad 65 \quad 70 \quad 70 \quad 74 \quad 79 \quad 81 \quad 85 \quad 90$$

To find the lower and upper quartiles, Q_1 and Q_3, of this batch, first find $\frac{1}{4}(n+1) = 2\frac{3}{4}$ and $\frac{3}{4}(n+1) = 8\frac{1}{4}$.

The lower quartile Q_1 is the number three quarters of the way from $x_{(2)}$ to $x_{(3)}$. These values are 60 and 65. The difference between them is $65 - 60 = 5$, and three quarters of that difference is $\frac{3}{4} \times 5 = 3.75$. Therefore Q_1 is 3.75 larger than 60, so it is 63.75. As with the median, in this module we will generally round the quartiles to the accuracy of the original data, so in this case we round to the nearest whole number, 64. In symbols, $Q_1 = 60 + \frac{3}{4}(65 - 60) = 63.75 \simeq 64$.

The upper quartile Q_3 is the number one quarter of the way from $x_{(8)}$ to $x_{(9)}$. These values are 81 and 85. The difference between them is $85 - 81 = 4$, and one quarter of that difference is $\frac{1}{4} \times 4 = 1$. Therefore Q_3 is 1 larger than 81, so it is 82. (No rounding necessary this time.) In symbols, $Q_3 = 81 + \frac{1}{4}(85 - 81) = 82$.

Example 15 is the subject of Screencast 3 for Unit 2 (see the module website).

Activity 10 Finding more quartiles

(a) Find the lower and upper quartiles of the batch of 15 coffee prices in Figure 14. (This batch of coffee prices was first introduced in Table 1 of Subsection 1.1.)

26	8 8 8 8 9
27	5 9
28	
29	5 5 5 5 9
30	5
31	5
32	
33	
34	
35	
36	9

$n = 15$ 26 | 8 represents 268 pence

Figure 14 Stemplot of 15 coffee prices

(b) Find the lower and upper quartiles of the batch of 14 gas prices in Figure 15. (This batch of gas prices was first introduced in Table 3 of Subsection 1.2.)

374	0 0 3
375	
376	0 7
377	6
378	4
379	5 6
380	1 1 4 5
381	8

$n = 14$ 374 | 0 represents 3.740p per kWh

Figure 15 Stemplot of 14 gas prices

A measure of spread

Now we can define a new measure of spread based entirely on the lower and upper quartiles.

The interquartile range

The interquartile range (sometimes abbreviated to **IQR**) is the distance between the lower and upper quartiles:

$$\text{IQR} = Q_3 - Q_1.$$

Note that this value is independent of the sizes of E_U and E_L.

Example 16 The prices of small televisions, yet again!

For the batch of 20 television prices in Example 14,

$$\begin{aligned}\text{IQR} &= Q_3 - Q_1 \\ &= 180 - 130 \\ &= 50.\end{aligned}$$

So the interquartile range is \$50.

Activity 11 Coffee prices again

Calculate both the range and the interquartile range of the batch of 15 coffee prices, last seen in Figure 14.

Activity 12 Interquartile range of gas prices

In Activity 10(b) you found the quartiles of the 14 gas prices from Activity 2 (Subsection 1.2). Find the interquartile range.

You may be wondering why you are being asked to learn a new measure of spread when you already know the range. As a measure of spread, the range $(E_U - E_L)$ is not very satisfactory because it is not resistant to the effects of unrepresentative extreme values. The interquartile range, by contrast, is a highly resistant measure of spread (because it is not sensitive to the effects of values lying outside the middle 50% of the batch) and it is generally the preferred choice.

Resistant measures were explained in Subsection 1.4.

Example 17 Comparing the resistance of the range and the IQR

Suppose the price of the most expensive jar of coffee is reduced from 369p to 325p. How does this affect the range and the interquartile range of the batch of coffee prices in Figure 14?

The new range is

$$E_U - E_L = 325\text{p} - 268\text{p} = 57\text{p},$$

a lot less than the original value of 101p (found in Activity 11). The interquartile range is unchanged.

3.3 The five-figure summary and boxplots

As well as giving us a new measure of spread – the interquartile range – the quartiles are important figures in themselves. Our $\wedge\!\wedge$-shaped diagram, Figure 16, gives five important points which help to summarise the shape of a distribution: the **median**, the **two quartiles** and the **two extremes**.

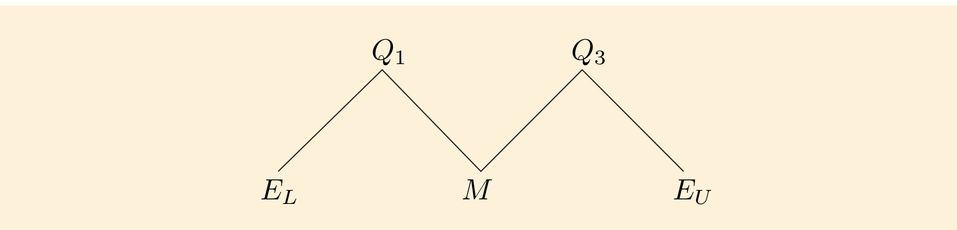


Figure 16 Values in a five-figure summary

These are conveniently displayed in the following form, called the **five-figure summary** of the batch.

Five-figure summary

		M	
n	Q_1		Q_3
	E_L		E_U

n batch size
M median
Q_1 lower quartile
Q_3 upper quartile
E_L lower extreme
E_U upper extreme

Example 18 Five-figure summary for television price data

You last saw these data in Figure 13.

For the television price data, we have $n = 20$, $M = 150$, $Q_1 = 130$, $Q_3 = 180$, $E_L = 90$ and $E_U = 270$.

Therefore, the five-figure summary of this batch is

		150	
$n = 20$	130		180
	90		270

This diagram contains the following information about the batch of prices.

- The general level of prices, as measured by the median, is $150.
- The individual prices vary from $90 to $270.
- About 25% of the prices were less than $130.
- About 25% of the prices were more than $180.
- About 50% of the prices were between $130 and $180.

We hope you agree that the five-figure summary is quite an efficient way of presenting a summary of a batch of data.

The five values in a five-figure summary can be very effectively presented in a special diagram called a **boxplot**. For the 14 gas prices (Figure 15) the diagram looks like Figure 17.

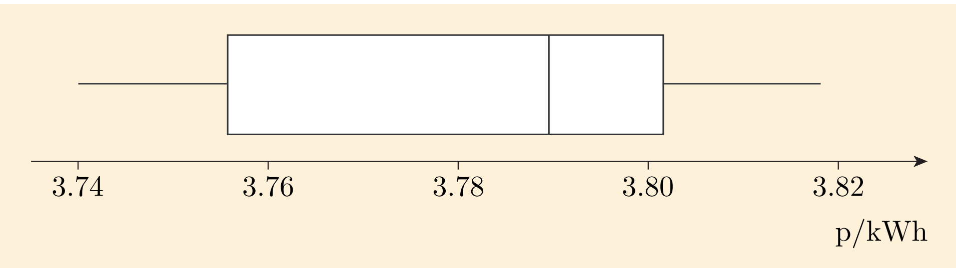


Figure 17 Boxplot of batch of 14 gas prices

The central feature of this diagram is a *box* – hence the name *box*plot. The box extends from the *lower quartile* (at the left-hand edge of the box) to the *upper quartile* (the right-hand edge). This part of the diagram contains 50% of the values in the batch. The length of this box is thus the *interquartile range*.

Outside the box are two *whiskers*. (Boxplots are sometimes called *box-and-whisker diagrams*.) In many cases, such as in Figure 17, the whiskers extend all the way out to the extremes. Each whisker then covers the end 25% of the batch and the distance between the two whisker-ends is then the *range*. (You will see examples later where the whiskers do not go right out to the extremes.)

So far we have dealt with four figures from the five-figure summary: the two quartiles and the two extremes. The remaining figure is perhaps the most important: it is the *median*, whose position is shown by putting a vertical line through the box.

Thus a boxplot shows clearly the division of the data into four parts: the two whiskers and the two sections of the box; these are the four parts of the /\/\-shaped diagram and each contains (approximately) 25% of values in the batch (see Figure 18).

John W. Tukey (1915–2000), inventor of the five-figure summary and boxplot

John Tukey was a prominent and prolific US statistician, based at Princeton University and Bell Laboratories. As well as working in some very technical areas, he was a great promoter of simple ways of picturing and summarising data, and invented both the five-figure summary and the boxplot (except that he called them the 'five-number summary' and the 'box-and-whisker plot').

He had what has been described as an 'unusual' lecturing style. The statistician Peter McCullagh describes a lecture he gave at Imperial College, London in 1977:

> Tukey ambled to the podium, a great bear of a man dressed in baggy pants and a black knitted shirt. These might once have been a matching pair, but the vintage was such that it was hard to tell. . . . The words came . . . , not many, like overweight parcels, delivered at a slow unfaltering pace. . . . Tukey turned to face the audience . . . . 'Comments, queries, suggestions?' he asked . . . . As he waited for a response, he clambered onto the podium and manoeuvred until he was sitting cross-legged facing the audience. . . . We in the audience sat like spectators at the zoo waiting for the great bear to move or say something. But the great bear appeared to be doing the same thing, and the feeling was not comfortable. . . . After a long while, . . . he extracted from his pocket a bag of dried prunes and proceeded to eat them in silence, one by one. The war of nerves continued . . . four prunes, five prunes. . . . How many prunes would it take to end the silence?
>
> (Source: McCullagh, P. (2003) 'John Wilder Tukey', *Biographical Memoirs of Fellows of the Royal Society*, vol. 49, pp. 537–555.)

John Tukey teaching at Princeton University

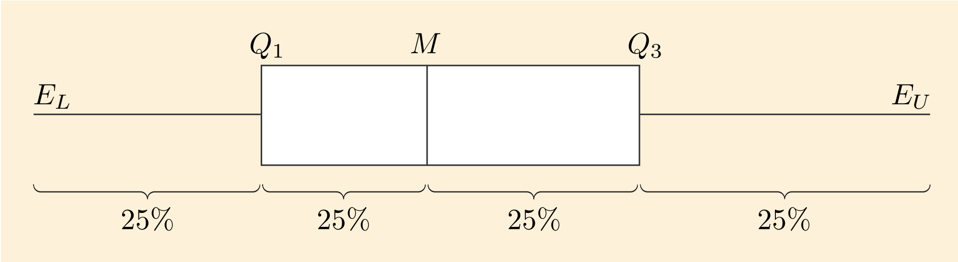


Figure 18 A standard boxplot with annotation

A typical boxplot looks something like Figure 18 because in most batches of data the values are more densely packed in the middle of the batch and are less densely packed in the extremes. This means that each whisker is usually longer than half the length of the box. This is illustrated again in the next example.

Example 19 Boxplot for the prices of small televisions

The boxplot for the batch of 20 television prices (last worked with in Example 18) is shown in Figure 19.

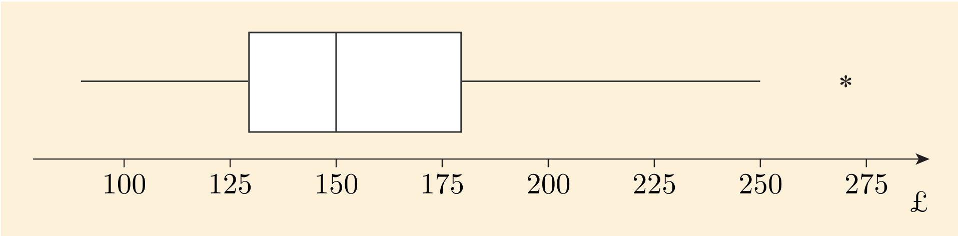


Figure 19 Boxplot of batch of 20 television prices

You can see that each whisker is longer than half the length of the box.

However, this boxplot has a new feature. The whisker on the left goes right down to the lower extreme. But the whisker on the right does not go right to the upper extreme. The highest extreme data value, 270, which might potentially be regarded as an outlier, is marked separately with a star. Then the whisker extends only to cover the data values that are not extreme enough to be regarded as potential outliers. The highest of these values is 250.

In Unit 3, you will learn in detail how to draw a boxplot. This includes a rule to decide which data values (if any) can be regarded as potential outliers that are plotted separately on the diagram.

Example 19 is the subject of Screencast 4 for Unit 2 (see the module website).

One important use of boxplots is to picture and describe the overall shape of a batch of data.

Use of boxplots will also be covered in Unit 3.

Example 20 Skew televisions

The stemplot of small television prices, last seen in Figure 13 (Subsection 3.2), shows a lack of symmetry. Since the higher values are more spread out than the lower values, the data are right-skew.

Skewness and symmetry were discussed in Subsection 5.2 of Unit 1.

The boxplot of these data, given in Figure 19, also shows this right-skew fairly clearly. In the box, the *right*-hand part (corresponding to higher prices) is rather longer than the left-hand part, and the *right*-hand whisker is longer than the left-hand whisker.

Activity 13 Skew gas prices?

A stemplot of the gas price data from Activity 2 (Subsection 1.2) is shown, yet again, in Figure 20.

374	0 0 3
375	
376	0 7
377	6
378	4
379	5 6
380	1 1 4 5
381	8

$n = 14$ 374 | 0 represents 3.740p per kWh

Figure 20 Stemplot of 14 gas prices

(a) Prepare a five-figure summary of the batch.

(b) Figure 21 shows the boxplot of these data that you have already seen in Figure 17. What do the stemplot and boxplot tell us about the symmetry and/or skewness of the batch?

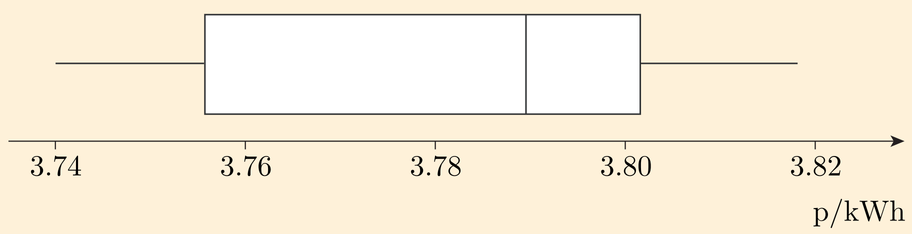


Figure 21 Boxplot of batch of 14 gas prices

Example 21 Camera prices: skew or not?

In Example 20 and Activity 13 you saw how boxplots look for batches of data that are right-skew or left-skew. What happens in a batch that is more symmetrical?

For the small batch of camera prices from Table 2 (Subsection 1.2), a (stretched) stemplot is shown in Figure 22.

5	3
5	
6	0
6	5
7	0 0 4
7	9
8	1
8	5
9	0

$n = 10$ 5 | 3 represents £53

Figure 22 Stemplot of ten camera prices

The stemplot looks reasonably symmetric.

A boxplot of the data, Figure 23, confirms the impression of symmetry. The two parts of the box are roughly equal in length, and the two whiskers are also roughly equal in length.

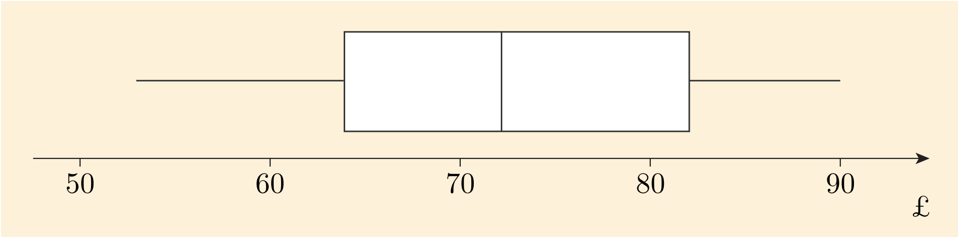


Figure 23 Boxplot of batch of ten camera prices

You have now spent quite a lot of time looking at various ways of investigating prices and, in particular, at methods of measuring the location and spread of the prices of particular commodities.

In order to begin to answer our question, *Are people getting better or worse off?*, we need to know not just location (and spread) of prices but also how these prices are *changing* from year to year. That is the subject of the rest of this unit.

Exercises on Section 3

Exercise 6 Finding quartiles and the interquartile range

(a) For the arithmetic scores in Exercise 1 (Section 1), find the quartiles and calculate the interquartile range. The stemplot of the scores is given below.

0	7
1	5
2	
3	3 5
4	2 2 3
5	5 8
6	4 6 8
7	1 1 6 8 9
8	0 1 1 3 4 5 5 6 9
9	1 1 3 5 9
10	0 0

$n = 33$ 0 | 7 represents a score of 7%

(b) For the television prices in Exercise 1, find the quartiles and calculate the interquartile range. The table of prices is given below.

170	180	190	200	220	229	230	230	230
230	250	269	269	270	279	299	300	300
315	320	349	350	400	429	649	699	

Exercise 7 Some five-figure summaries

Prepare a five-figure summary for each of the two batches from Exercise 1.

(a) For the arithmetic scores, the median is 79% (found in Exercise 1), and you found the quartiles and interquartile range in Exercise 6.

(b) For the television prices, the median is $\$270$ (found in Exercise 1), and you found the quartiles and interquartile range in Exercise 6.

Exercise 8 Boxplots and the shape of distributions

Boxplots of the two batches used in Exercises 1, 6 and 7 are shown in Figures 24 and 25. On the basis of these diagrams, comment on the symmetry and/or skewness of these data.

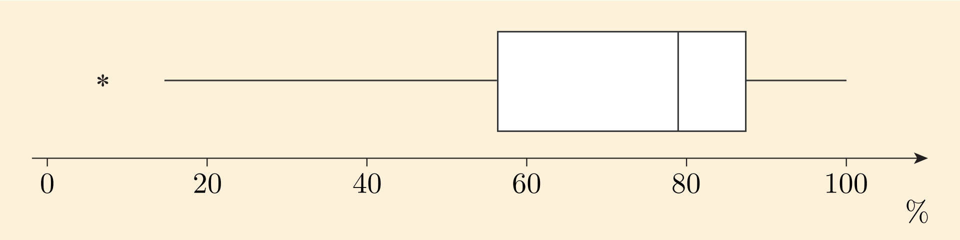


Figure 24 Boxplot of batch of 33 arithmetic scores

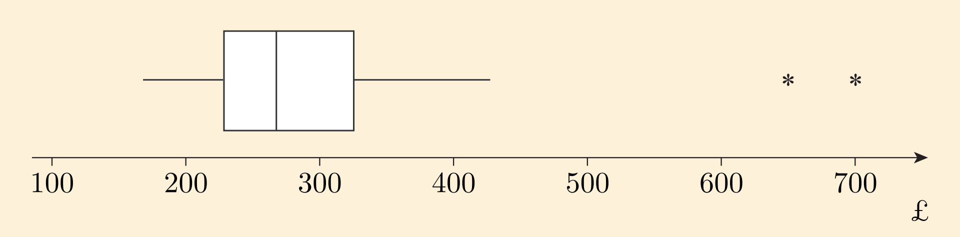


Figure 25 Boxplot of batch of 26 television prices

4 A simple chained price index

You have already seen that it is not a simple task to measure the price of even a *single* commodity at a *fixed* time and place. Measuring the change in price of a single commodity from one year to the next will be even more complicated but, as was said in Subsection 1.1, to answer our question it is necessary to measure the *changes* in the prices of the *whole range* of goods and services which people use. Moreover, since we wish to know how all the different changes in the prices of these goods and services affect people, we need to take into account those people's consumption patterns. For example, a large increase in the price of high-quality caviar will not affect most people's budgets since most households' shopping lists do not include this commodity!

This makes the task of measuring price changes and examining how they affect us seem exceedingly difficult; but such a task is carried out in the UK regularly each month, organised by the Office for National Statistics. (Most of the prices are actually collected by a market research company under contract to the Office for National Statistics.) The results of their data collection and subsequent calculations are summarised in two measures called the Consumer Prices Index (CPI) and the Retail Prices Index (RPI).

'index'

These indices do not measure prices. Each is an index of price *changes* over time, and one or both of these indices are commonly used when people make comparisons about the cost of living. As you will see in Unit 3, they are highly relevant measures for those engaged in wage bargaining.

The RPI and the CPI are both ‘chained’ in the sense that the index value for each year is linked to the year before. The very first link in the chain is called the *base year* and it is given an index value of 100.

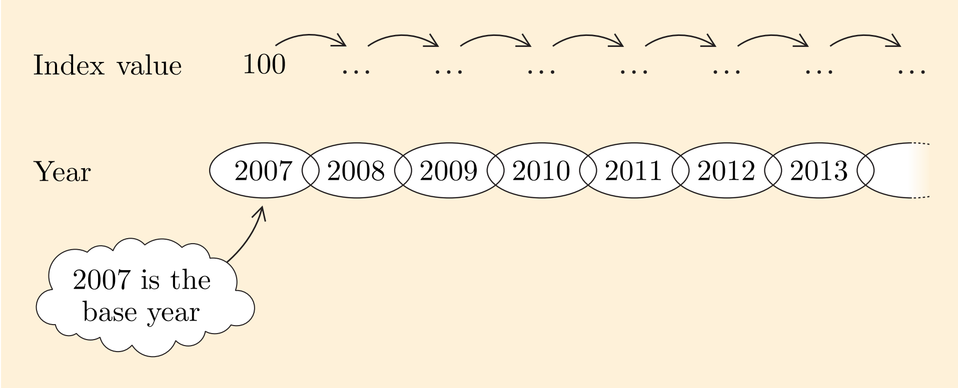


Figure 26 A chained index

The original Mr. Gradgrind, from an 1870s illustration to Charles Dickens’ *Hard Times*, first published in 1854. Dickens had Gradgrind describe himself as, ‘A man of realities. A man of facts and calculations.’ But don’t let that put you off.

4.1 A two-commodity price index

Section 5 includes an outline of how the information used to calculate the official UK price indices is collected, and describes how the indices are calculated. To introduce ideas, in this section we describe a very much simpler example of a price index calculation. It uses exactly the same basic method of calculation as the actual Retail Prices Index. (Not every index is calculated in this way, as you will see in Unit 3 with the Average Weekly Earnings statistic.)

The context is a mythical computing company, Gradgrind Ltd, whose organisation and exploits will be used occasionally in this and later units to illustrate various points.

Gradgrind Ltd uses both gas and electricity in its operations. Table 7 shows the price they paid for each fuel in 2007 and 2008. The prices are shown in *$* per megawatt hour (MWh). (It is more usual, in the UK, for prices to be quoted in pence per kilowatt hour (p/kWh). Here, *$*/MWh have been used simply to make some of the later calculations a little more straightforward. Because there are 100 pence in *$*1 and 1000 megawatts in a kilowatt, *$*10/MWh is exactly the same price as 1p/kWh – so Gradgrind’s gas price in 2007, for instance, was 2.4p/kWh.)

Table 7 Gradgrind’s energy prices in 2007 and 2008

	2007	2008
Gas ($/MWh)	24	29
Electricity ($/MWh)	76	87

If we were interested in looking at the change in price of just *one* of these fuels, say gas, things would be relatively straightforward. For instance, it might well be appropriate to look at the increase in price as a percentage of the price in 2007.

Activity 14 Gradgrind's gas price increase

Work out the increase in Gradgrind's gas price between 2007 and 2008 as a percentage of the 2007 price.

So we could say that, for this company at least, gas has gone up by 20.8%. In other words, for every $1 they spent on gas in 2007, they would have spent $1.208 in 2008 if they had bought the same amount of gas in each year. Or putting it another way, for every 100 units of money (pence, pounds, whatever) they spent in 2007, they would have spent 120.8 units of money in 2008 if they had bought the same amount. So a way of representing this price change would have been to define an *index* for the gas price such that it takes the value 100 for 2007, and 120.8 for 2008.

Notice that the value of the gas price index for 2008 could be calculated as

$$(\text{value of the index in 2007, which is taken as } 100) \times \frac{\text{gas price in 2008}}{\text{gas price in 2007}}.$$

That is, the value of the index in one year is the value of the index in the previous year multiplied by a *price ratio*, in this case the *gas price ratio for 2008 relative to 2007*. This ratio, as a number, is 1.208.

But Gradgrind did not only use gas, they used electricity as well, and the aim here is to find a representation of their overall fuel price change, not just the change in gas prices.

An *electricity price ratio for 2008 relative to 2007* can be worked out, like the gas price ratio. It is $\frac{87}{76} \simeq 1.145$.

Activity 15 Gradgrind's electricity price index

Use the electricity price ratio above to find the increase in Gradgrind's electricity price between 2007 and 2008 as a percentage of the 2007 price. What would the 2008 value be for a price index of Gradgrind's electricity price alone, calculated in the same way as the gas price index (with 2007 as the base year)?

But this has got us no further in finding a price index that simultaneously covers *both* fuels.

One possibility might be to look at how Gradgrind's total expenditure on these two fuels changed from 2007 to 2008. The expenditures are given in Table 8.

Table 8 Gradgrind's energy expenditure ($) in 2007 and 2008

	2007	2008
Gas	9 298	8 145
Electricity	3 205	2 991
Total	12 503	11 136

This seems not to have helped. The total expenditure went down, but you have already seen that the prices of both gas and electricity went up.

Activity 16 How much fuel did Gradgrind use?

Use the data in Tables 7 and 8 to find the quantity of each fuel that Gradgrind used in 2007 and 2008 (in MWh). Hence explain why the energy expenditure fell.

Remember the aim is to produce a measure of price *changes*. So looking at expenditure changes does not do the right thing, since expenditure depends on the amount of fuel consumed as well as the price.

One possibility might be as follows. We could work out how much Gradgrind *would have* spent on fuel in 2008 if the consumptions of both fuels had not changed from 2007. That would remove the effect of any changes in consumption. Then we could calculate an overall energy price ratio for 2008 relative to 2007 by dividing the total expenditure on energy for 2008 (using the 2007 consumption figures) by the total expenditure on energy for 2007 (again using the 2007 consumption figures).

You should have found, in Activity 16, that the quantities of gas and electricity consumed in 2007 were, respectively, 387.4 MWh and 42.2 MWh. To buy those quantities at 2008 prices would have cost (in \$): $29 \times 387.4 = 11\,234.6$ for the gas and $87 \times 42.2 = 3671.4$ for the electricity, giving a total expenditure of

$$\$(11\,234.6 + 3671.4) = \$14\,906.0.$$

So a reasonable overall energy price ratio for 2008 relative to 2007 can be found by dividing this total by the 2007 total expenditure, again calculated using the 2007 consumptions. The appropriate figure for 2007 is just the actual total expenditure, which (in \$) was $9298 + 3205 = 12\,503$ (see Table 8). This gives an overall energy price ratio for 2008 relative to 2007 as

$$\frac{14\,906.0}{12\,503} \simeq 1.192.$$

Now we have an appropriate price ratio, the Gradgrind energy price index can be set as 100 for the base year, 2007, and the value of the 2008 index is found by multiplying the 2007 index value by the price ratio:

$$\text{2008 index} = 100 \times 1.192 = 119.2.$$

This is indeed how a chained index of this kind is calculated – but the calculations are rather messy. You might be wondering whether it would be simpler to calculate the overall energy price ratio as a weighted mean of the two price ratios for the two fuels, in much the same way that weighted means were used to combine prices in Section 2. If you did think this, you would be right – and furthermore, the resulting overall energy price ratio is exactly the same as has just been found, if we make the right choice of weights. The overall energy price ratio for 2008 relative to 2007 is just a weighted mean of the two price ratios for gas and electricity, with the 2007 expenditures as weights.

Just to show it really does come to the same thing, let us see how it works with the numbers, using the formula for weighted means in Subsection 2.3.

	Price ratio (2008 relative to 2007)	Weight (2007 expenditure)
Gas	1.208	9298
Electricity	1.145	3205

The weighted average of these price ratios is

$$\frac{(1.208 \times 9298) + (1.145 \times 3205)}{9298 + 3205} = \frac{14\,901.709}{12\,503} \simeq 1.192,$$

giving the same value for the overall energy price ratio for 2008 relative to 2007 as we found earlier. (And this is not some sort of fluke that applies only to these particular numbers; it can be shown mathematically that it always works.)

Activity 17 Gradgrind's energy price ratio for 2009 relative to 2008

Table 9 Gradgrind's energy prices and expenditures for 2008 and 2009

	2008	2009
Gas price ($/MWh)	29	30
Gas expenditure ($)	8 145	23 733
Electricity price ($/MWh)	87	98
Electricity expenditure ($)	2 991	2 275

(a) Using the data in Table 9, calculate the price ratios for gas and for electricity, in each case for 2009 relative to 2008.

(b) With the 2008 expenditures as weights, use your answers to part (a) to calculate the overall energy price ratio for 2009 relative to 2008.

(c) Now see what happens if you use the 2009 expenditures as weights to calculate the overall energy price ratio for 2009 relative to 2008. How do the results of the calculation differ from what you got in part (b)?

The reason that the price ratios you calculated in parts (b) and (c) in Activity 17 were so different is that Gradgrind's 'energy mix' changed a lot over the year. Compared with 2008, in 2009 they spent a great deal more on gas but less on electricity. The weighted mean of the gas and electricity price ratios is, in both cases, nearer the price ratio for gas than that for electricity – this is Rule 2 for weighted means – but it is even nearer the gas weighted mean when the 2009 expenditures are used. This is because the weight for gas is proportionally much greater than it is when the 2008 expenditures are used as weights.

This all shows that it *does* make a difference which expenditures are used as weights. In practice, it is much more common to use the expenditures from the earlier year – 2008 in this case – as weights. In some circumstances, though, there are good reasons for using the later year, or indeed some more complicated set of weights that depend on both expenditures. However, in this unit we shall use the expenditures from the earlier year to provide the weights, partly because that matches more closely what is done in calculating the official UK price indices.

Another possibility for weights would have been to continue to use the 2007 expenditures. These were used to find the overall energy price ratio for 2008 relative to 2007 and could be used for later years as well. Again, in some circumstances this would make sense, but here the pattern of Gradgrind's fuel expenditure has changed a lot over time, and weights should change in consequence. To continue to use the 2007 expenditures for all later years would mean that this change in the relative importance to Gradgrind of the two fuels would never be taken into account. Instead, to obtain the overall energy price

ratio from one year to the next, we use the fuel expenditures in the earlier year as weights, so each year the weights change.

That determines the choice of weights in forming an overall price ratio. Now, how is that used to find the energy price *index*? Here we simply continue the 'chaining' that started when finding the 2008 index: the 2009 index is found by multiplying the value of the index for the previous year, 2008, by the overall energy price ratio for 2009 relative to 2008. The value of the index for 2008 was calculated earlier as 119.2, and (using the weights from the previous year) the overall energy price ratio for 2009 relative to 2008 was found in Activity 17(b) as 1.059. So the value of Gradgrind's energy price index for 2009 is

$$119.2 \times 1.059 \simeq 126.2.$$

(So, in a particular kind of average way, Gradgrind's energy prices for 2009 have risen by 26.2% since the base year, 2007.)

In general, the value index for a particular year is found by multiplying the value of the index for the previous year by the overall energy price ratio for that year relative to the previous year. This is illustrated in Figure 27.

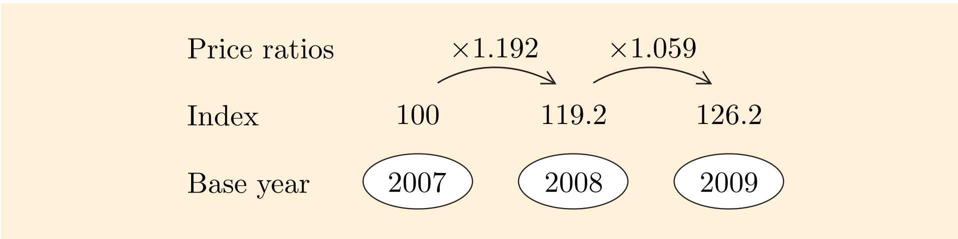


Figure 27 Determining a chained price index

In the process of *chaining*, the overall price ratio is calculated anew each year, looking back only at the previous year. The ratio is used to 'chain' to earlier years and hence determine the value of the index. This method of calculating a **chained price index** is summarised below. Although there were only two commodities (gas and electricity) in Gradgrind's index, this summary is not restricted to two commodities.

Procedure used to calculate a chained price index

1. For each year calculate the following.
   - The **price ratio** for each commodity covered by the index:
     $$\frac{\text{price that year}}{\text{price previous year}}.$$
   - The weighted mean of all these price ratios, using as weights the expenditure on each commodity in the previous year. This weighted mean is called the **all-commodities price ratio**.
2. For each year, the value of the index is

   value of index for previous year × all-commodities price ratio.

   The value of the index in the first year is set at 100; this date is the **base date** of the index.

Activity 18 Gradgrind's energy price index for 2010

Use the data in Table 10, and other necessary numbers from previous calculations, to calculate the value of Gradgrind's energy price index for 2010.

Table 10 Gradgrind's energy prices and expenditures for 2009 and 2010

	2009	2010
Gas price ($/MWh)	30	28
Gas expenditure ($)	23 733	23 969
Electricity price ($/MWh)	98	88
Electricity expenditure ($)	2 275	2 920

The Retail Prices Index (RPI), published by the UK Office for National Statistics, is calculated once a month rather than once a year, but the method used is basically that outlined above, though with far more than two commodities. The process of finding the weights in the Retail Prices Index is also more complicated, because it involves taking into account the expenditures of millions of people as measured in a major survey. However, the principles are the same as for Gradgrind. The calculation each January follows exactly this method. In the other 11 months of the year, the calculation is very similar but uses only the increases in prices since the previous January. In the next section, you will learn more about how all this works.

See Subsection 5.2 for the details of these calculations.

Exercise on Section 4

Exercise 9 Gradgrind's energy price index for 2011

Use the data in Table 11, and the fact that Gradgrind's energy price index for 2010 was 117.4 (as found in Activity 18), to calculate the value of Gradgrind's energy price index for 2011.

Table 11 Gradgrind's energy prices and expenditures for 2010 and 2011

	2010	2011
Gas price ($/MWh)	28	30
Gas expenditure ($)	23 969	24 282
Electricity price ($/MWh)	88	86
Electricity expenditure ($)	2 920	3 117

5 The UK government price indices

'The huge squeeze on Brits was laid bare today as figures showed inflation has soared to a 20-year high.' (*The Sun*, 18 October 2011)

'Overall, prices in the economy rose 0.6% on the month from August.' (*Guardian*, 18 October 2011)

'Inflation in the UK continued to fall in February, thanks largely to lower gas and electricity bills.' (BBC News website, 20 March 2012)

'UK inflation rises more than expected.'
(*Daily Telegraph*, 16 August 2011)

How often have you read or heard statements like these in the media? Have you ever wondered how 'inflation' is measured, or precisely what is meant by a statement such as 'prices rose by 0.6%'? In Subsection 5.3, you will see that 'rates of inflation' are often calculated in the UK using an index of prices paid by consumers, the Consumer Prices Index (CPI), or another slightly different index, the Retail Prices Index (RPI). These indices may be used to calculate the percentage by which prices in general have risen over any given period, and (roughly speaking) this is what is meant by inflation. But what exactly do these price indices measure, and how are they calculated? These are the questions that are addressed in this section.

5.1 What are the CPI and RPI?

The CPI and the RPI are the main measures used in the UK to record changes in the level of the prices most people pay for the goods and services they buy. The RPI is intended to reflect the average spending pattern of the great majority of private households. Only two classes of private households are excluded, on the grounds that their spending patterns differ greatly from those of the others: pensioner households and high-income households. The CPI, however, has a wider remit – it is intended to reflect the spending of *all* UK residents, and also covers some costs incurred by foreign visitors to the UK.

The CPI and RPI are calculated in a similar way to the price index for Gradgrind Ltd's energy in Section 4. However, they are calculated once a month rather than just once a year, and are based on a very large '**basket of goods**'. The contents of the basket and the weights assigned to the items in the basket are updated annually to reflect changes in spending patterns (as was the case with Gradgrind's index for energy prices), and the index is 'chained' to previous values. However, once decided on at the beginning of the year, the contents of the basket and their weights remain fixed throughout the year.

For the RPI, the price ratio for the basket each month is calculated relative to the previous January. Then the value of the index is obtained by multiplying the value of the index for the previous January by this price ratio. For example,

$$\begin{aligned}\text{RPI for Nov. 2011} &= \text{RPI for Jan. 2011}\\ &\quad \times \text{ (price ratio for Nov. 2011 relative to Jan. 2011).}\end{aligned}$$

The CPI works in much the same way, except that price ratios are calculated relative to the previous December. So, for example,

$$\begin{aligned}\text{CPI for Nov. 2011} &= \text{CPI for Dec. 2010}\\ &\quad \times \text{ (price ratio for Nov. 2011 relative to Dec. 2010).}\end{aligned}$$

Since these price indices are calculated from price ratios, they measure price changes in terms of the *ratio* of the overall level of prices in a given month to the overall level of prices at an earlier date. In practice, data on most prices are collected on a particular day near the middle of the month; the values of the RPI and CPI calculated using these data are referred to simply as the values of the RPI and CPI for the month. For example, the RPI took the value 239.9 in February 2012. This value measures the ratio of the overall level of prices in February 2012 to the overall level of prices on a date at which the index was fixed at its starting value of 100. This date, called a *base date*, is 13 January 1987 (at the time of writing). Thus the general level of prices in February 2012, as measured by the RPI, was $239.9/100 = 2.399$ times the general level of prices in January 1987. The base date has *no* significance other than to act as a reference point. (The CPI base date is 2005 and this refers to the average level of prices throughout 2005, not to a specific date in 2005.)

The RPI and CPI are each based on a very large 'basket' of goods and services. (The two baskets are similar, but not exactly the same.) Each contains around 700 items including most of the usual things people buy: food, clothes, fuel, household goods, housing, transport, services, and so on. Each basket is an 'average' basket for a broad range of households. The items in the baskets are often grouped into broader categories. For the RPI, the five fundamental groups are: 'Food and catering', 'Alcohol and tobacco', 'Housing and household expenditure', 'Personal expenditure' and 'Travel and leisure'. These groups are divided into 14 more detailed *subgroups* (which are further divided into *sections*), as shown in Figure 28.

The items in the CPI basket are divided into 12 broad groupings called *divisions*, which are further subdivided.

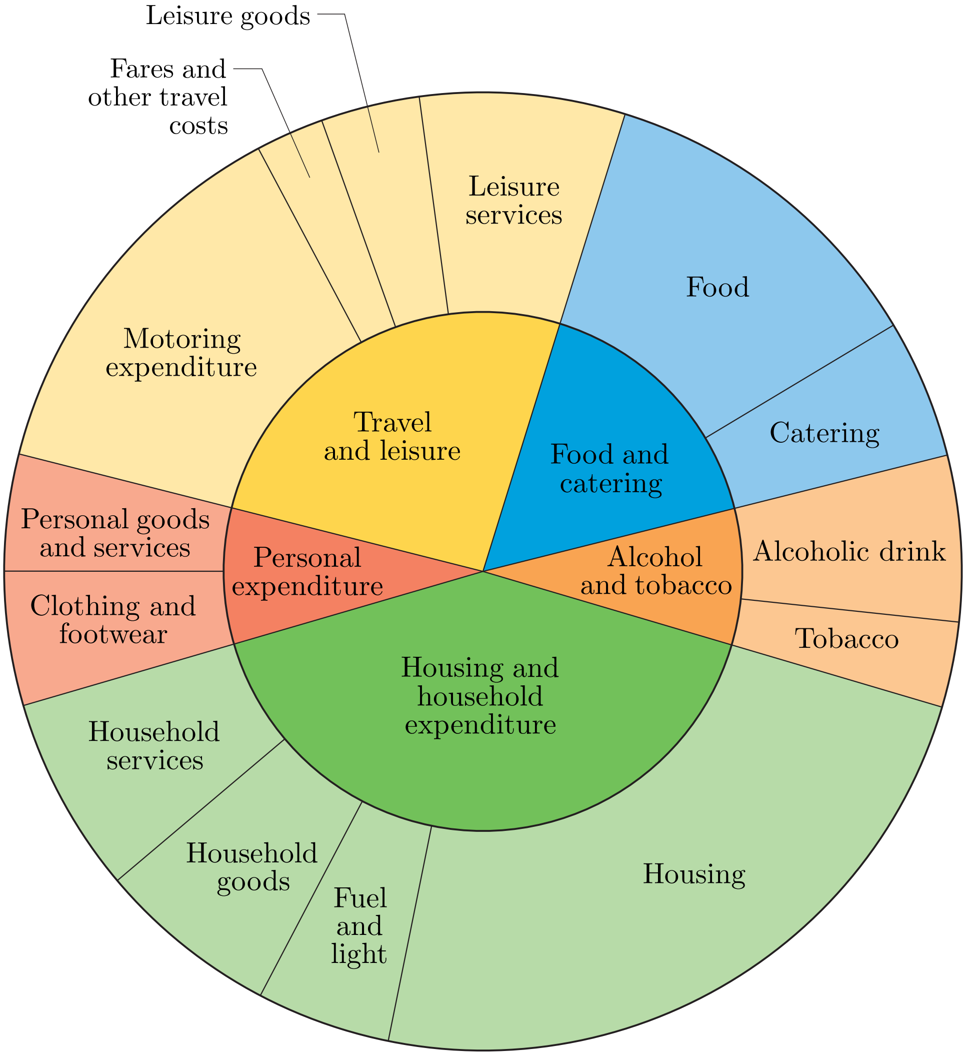


Figure 28 Structure of the RPI in 2012 (based on data from the Office for National Statistics)

The inner circle shows the five groups, and the outer ring shows the 14 subgroups. Notice that in the inner circle the sector labelled 'Food and catering' has been drawn almost twice as large (as measured by area) as that labelled 'Alcohol and tobacco'. This reflects the fact that the typical household spends nearly twice as much on food and catering as on alcohol and tobacco. The weight of an item or group reflects how much money is spent on it. So the weight of the 'Food and catering' group is almost twice that of 'Alcohol and tobacco'.

The outer ring represents the same total expenditure as the inner circle, but in more detail. For example, in the outer ring the area labelled 'Food' (which mostly consists of food bought for use in the home) is more than twice as large as that

labelled ‘Catering’ (which includes meals in restaurants and canteens, and take-away meals and snacks), reflecting the fact that the typical household spends more than twice as much on food as on catering; the weight of the subgroup ‘Food’ is more than double the weight of the subgroup ‘Catering’. The chart gives a good indication of average spending patterns in the UK in the early 21st century.

Activity 19 The expenditure of a typical household

(a) Using Figure 28, estimate roughly what fraction of the expenditure of a typical household is on each of the following groups and subgroups:

- Personal expenditure
- Housing and household expenditure
- Housing

(b) Suppose that a household spends a total of £540 per week on goods and services that are covered by the RPI. Use your answers to part (a) to estimate very approximately how much is spent each week on each of the groups and subgroups in part (a).

To ensure that the basket of goods for the index reflects the proportion of average spending devoted to different types of goods and services, it is necessary to find out how people actually spend their money. The Living Costs and Food Survey (LCF) records the spending reported by a sample of 5000 households spread throughout the UK. Data from the LCF are used to calculate the weights of most of the items included in the RPI basket. Since 1962, the weights have been revised each year, so that the index is always based on a basket of goods and services that is as up to date as possible. Because of this regular weight revision, the index is chained (as was the Gradgrind Ltd index).

(Most of the weights for the CPI come from a different source, the UK National Accounts, though in turn this source is partly based on data from the LCF. Again, the weights are revised each year.)

The weight of a group or subgroup directly depends on the average expenditure of households on that item. In Subsection 2.1, you saw that it is only the *relative* size of the weights that affects the value of the weighted mean – this is Rule 1 for weighted means. So instead of using the average expenditure of an item as its weight, the expenditure figures for the items can all be multiplied by the same factor to produce a new, more convenient, set of weights. For the RPI, this factor is chosen so that the sum of the weights is 1000. Table 12 shows the 2012 weights used in the RPI for the groups and subgroups. Notice that each group weight is obtained by summing the weights for its subgroups.

Table 12 2012 RPI weights

Group	Subgroup	Weight	Group weight
Food and catering	Food	114	
	Catering	47	161
Alcohol and tobacco	Alcoholic drink	56	
	Tobacco	29	85
Housing and household expenditure	Housing	237	
	Fuel and light	46	
	Household goods	62	
	Household services	67	412
Personal expenditure	Clothing and footwear	45	
	Personal goods and services	39	84
Travel and leisure	Motoring expenditure	131	
	Fares and other travel costs	23	
	Leisure goods	33	
	Leisure services	71	258
All items (i.e. the sum of the weights)			1000

The following checklist provided contains the major categories of goods and services included in the RPI. In the next activity, you will be asked to complete the last three columns of this checklist to make rough estimates of your household's group weights.

A checklist for one household's average monthly expenditure

	Expenditure and weights			Your expenditure and weights		
	Expenditure 2012 (£)	Group totals (£)	Group weights	Expenditure 2012 (£)	Group totals (£)	Group weights
Food and catering						
– at home	370					
– canteens, snacks and take-aways	80					
– restaurant meals	20					
		470	266		____	____
Alcohol and tobacco						
– alcoholic drink	8					
– cigarettes and tobacco	0					
		8	5		____	____
Housing and household expenditure						
– mortgage interest/rent	82					
– council tax	95					
– water charges	47					
– house insurance	29					
– repairs/maintenance/DIY	40					
– gas/electricity/coal/oil bills	210					
– household goods (furniture appliances, consumables, etc.)	70					
– telephone and internet bills	20					
– school and university fees	0					
– pet care	0					
		593	336		____	____
Personal expenditure						
– clothing and footwear	45					
– other (hairdressing, chemists' goods, etc.)	10					
		55	31		____	____
Travel and leisure						
– motoring (purchase, maintenance, petrol, tax, insurance)	210					
– fares	200					
– books, newspapers, magazines	80					
– audio-visual equipment, CDs, etc.	15					
– toys, photographic and sports goods	3					
– TV purchase/rental, licence	0					
– cinema, theatre, etc.	30					
– holidays	100					
		638	362		____	____
		1764	1000		____	____

The figures already in the checklist were completed for a two-person household. Some of the figures were accurate, others were necessarily very rough estimates. Nevertheless, the household's weights give a reasonable indication of the proportion of the household's expenditure (in 2012) on the five main groups used in the RPI.

The total expenditure was $1764. So the group weights were calculated by multiplying all the group total expenditures by a constant factor of 1000/1764, to ensure the weights sum to 1000. The weight for 'Food and catering', for example, is

$$470 \times \frac{1000}{1764} \simeq 266.$$

Another way to calculate this is to multiply the proportion of monthly expenditure spent on food and catering by 1000. The proportion is

$$\frac{470}{1764} \simeq 0.266.$$

Since the total weight is 1000, the weight for 'Food and catering' is

$$0.266 \times 1000 = 266.$$

Notice that the group weights for this particular household differ quite considerably from those used in the RPI in 2012 (see Table 12). For instance, a much greater proportion of expenditure is on 'Food and catering' and a much smaller proportion is spent on 'Alcohol and tobacco'.

Activity 20 Your own household's expenditure

Make rough estimates of your own household's expenditure last year and complete the final columns of the checklist above. For some categories, you may find it easier just to make a rough estimate of, say, your annual expenditure and then divide by 12. If you have no idea at all for a category, then use the corresponding figure in the checklist as a starting point for your own expenditure and adjust it up or down depending on how you think you spend your money. One way of checking that your figures are sensible is to consider how the sum of the expenditures relates to your household's monthly income. Do not spend more than 15 minutes on estimating your expenditure; accurate figures are not needed.

Divide each group expenditure by your monthly expenditure total and then multiply by 1000 to calculate your household's group weights.

How do your household's weights compare with those used in the RPI in 2012?

5.2 Calculating the price indices

This subsection concentrates on how the RPI is calculated. Generally the CPI is calculated in a similar way, though some of the details differ. To measure price changes in general, it is sufficient to select a limited number of representative items to indicate the price movements of a broad range of similar items. For each section of the RPI, a number of *representative items* are selected for pricing. The selection is made at the beginning of the year and remains the same throughout the year. It is designed in such a way that the price movements of the representative items, when combined using a weighted mean, provide a good estimate of price movements in the section as a whole.

For example, in 2012 the representative items in the 'Bread' section (which is contained in the 'Food and catering' group) were: large white sliced loaf, large white unsliced loaf, large wholemeal loaf, bread rolls, garlic bread. Changes in the prices of these types of bread are assumed to be representative of changes in bread prices as a whole. Note that although the *price ratio* for bread is based on this sample of five types of bread, the calculation of the appropriate *weight* for bread is based on *all* kinds of bread. This weight is calculated using data collected in the Living Costs and Food Survey.

Collecting the data

The bulk of the data on price changes required to calculate the RPI is collected by staff of a market research company and forwarded to the Office for National Statistics for processing. Collecting the prices is a major operation: well over 100 000 prices are collected each month for around 560 different items. The prices being charged at a large range of shops and other outlets throughout the UK are mostly recorded on a predetermined Tuesday near the middle of the month. Prices for the remaining items, about 140 of them, are obtained from central sources because, for example, the prices of some items do not vary from one place to another.

One aim of the RPI is to make it possible to compare prices in any two months, and this involves calculating a value of the price index itself for every month.

Changing the representative items

The Office for National Statistics (ONS) updates the basket of goods every year, reflecting advancing technology, changing tastes and consumers' spending habits. The media often have fun writing about the way the list of representative items changes each year.

> In the 1950s, the mangle, crisps and dance hall admissions were added to the basket, with soap flakes among the items taken out.
>
> Two decades later, the cassette recorder and dried mashed potato made it in, with prunes being excluded.
>
> Then after the turn of the century, mobile phone handsets and fruit smoothies were included. The old fashioned staples of an evening at home – gin and slippers – were removed from the basket.
>
> So now, in 2012, it is the turn of tablet computers to be added to mark the growing popularity of this type of technology.

> That received the most coverage when it was added to the basket of goods, with the ONS highlighting this digital-age addition in its media releases.
>
> But those seafaring captains who once used the then unusual fruit as a symbol to show they were home and hosting might be astonished to find that centuries on, the pineapple has also been added to the inflation basket.
>
> Technically, the pineapple has been added to give more varied coverage in the basket of fruit and vegetables, the prices of which can be volatile.
>
> (Source: BBC News website, 14 March 2012)

So, calculating the RPI involves two kinds of data:

- the price data, collected every month
- the weights, representing expenditure patterns, updated once a year.

Once the price data have been collected each month, various checks, such as looking for unbelievable prices, are applied and corrections made if necessary. Checking data for obvious errors is an important part of any data analysis.

Then an averaging process is used to obtain a price ratio for each item that fairly reflects how the price of the item has changed across the country. The exact details are quite complicated and are not described here. (If you want more details, they are given in the *Consumer Price Indices Technical Manual*, available from the ONS website. *Consumer Price Indices: A brief guide* is also available from the same website.) For each item, a price ratio is calculated that compares its price with the previous January. For instance, for November 2011, the resulting price ratio for an item is an average value of

$$\frac{\text{price in November 2011}}{\text{price in January 2011}}.$$

The next steps in the process combine these price ratios, using weighted means, to obtain 14 subgroup price ratios, and then the group price ratios for the five groups. Finally, the group price ratios are combined to give the **all-item price ratio**. This is the price ratio, relative to the previous January, for the 'basket' of goods and services as a whole that make up the RPI.

The all-item price ratio tells us how, on average, the RPI 'basket' compares in price with the previous January. The value of the RPI for a given month is found by the method described in Section 4, that is, by multiplying the value of the RPI for the previous January by the all-item price ratio for that month (relative to the previous January):

$$\begin{aligned}\text{RPI for month } x = &(\text{RPI for previous January})\\ &\times\ (\text{all-term price ratio for month } x)\end{aligned}$$

Thus, to calculate the RPI for November 2011, the final step is to multiply the value of the RPI in January 2011 by the all-item price ratio for November 2011.

Example 22 Calculating the RPI for November 2011

Here are the details of the last two stages of calculation of the RPI for November 2011, after the group price ratios have been calculated, relative to January 2011. The appropriate data are in Table 13.

Table 13 Calculating the all-item price ratio for November 2011

Group	Price ratio r	Weight w	Ratio × weight rw
Food and catering	1.030	165	169.950
Alcohol and tobacco	1.050	88	92.400
Housing and household expenditure	1.037	408	423.096
Personal expenditure	1.128	82	92.496
Travel and leisure	1.026	257	263.682
Sum		1000	1041.624

(Source: Office for National Statistics)

You may have noticed that the weights here do not exactly match those in Table 12. That is because the weights here are the 2011 weights, and those in Table 12 are the 2012 weights, and as has been explained, the weights are revised each year.

The all-item price ratio is a weighted average of the group price ratios given in the table. If the price ratios are denoted by the letter r, and the weights by w, then the weighted mean of the price ratios is the sum of the five values of rw divided by the sum of the five values of w. The formula, from Subsection 2.3, is

$$\begin{aligned}\text{all-item price ratio} &= \frac{\text{sum of products (price ratio} \times \text{weight)}}{\text{sum of weights}}\\ &= \frac{\sum rw}{\sum w}.\end{aligned}$$

The sums are given in Table 13. (The sum of the weights is 1000, because the RPI weights are chosen to add up to 1000.) Although Table 13 gives the individual rw values, there is no need for you to write down these individual products when finding a weighted mean (unless you are asked to do so). As mentioned previously, your calculator may enable you to calculate the weighted mean directly, or you may use its memory to store a running total of rw.

Now the all-item price ratio for November 2011 (relative to January 2011) can be calculated as

$$\frac{1041.624}{1000} = 1.041\,624.$$

This tells us that, on average, the RPI basket of goods cost 1.041 624 times as much in November 2011 as in January 2011.

The published value of the RPI for January 2011 was 229.0. So, using the formula,

$$\begin{aligned}\text{RPI for Nov. 2011} =&\text{RPI for Jan. 2011}\\ &\times \text{(all-item price ratio for Nov. 2011)}\\ =&229.0 \times 1.041\,624\\ =&238.531\,896 \simeq 238.5.\end{aligned}$$

The final result has been rounded to the same number of decimal places as the group price ratios. This matches the published value of the RPI for November 2011.

Example 22 is the subject of Screencast 5 for Unit 2 (see the module website).

The same 2011 weights were used to calculate the RPI for every month from February 2011 to January 2012 inclusive. For each of these months, the price ratios were calculated relative to January 2011, and the RPI was finally

calculated by multiplying the RPI for January 2011 by the all-item price ratio for the month in question. In February 2012, however, the process began again (as it does every February). A new set of weights, the 2012 weights, came into use. Price ratios were calculated relative to January 2012, and the RPI was found by multiplying the RPI value for January 2012 by the all-item price ratio. This procedure was used until January 2013, and so on.

The process of calculating the RPI can be summarised as follows.

Calculating the RPI

1. The data used are prices, collected monthly, and weights, based on the Living Costs and Food Survey, updated annually.
2. Each month, for each item, a price ratio is calculated, which gives the price of the item that month divided by its price the previous January.
3. Group price ratios are calculated from the price ratios using weighted means.
4. Weighted means are then used to calculate the all-item price ratio. Denoting the group price ratios by r and the group weights by w, the all-item price ratio is
$$\frac{\sum rw}{\sum w}.$$
5. The value of the RPI for that month is found by multiplying the value of the RPI for the previous January by the all-item price ratio:
$$\begin{aligned}\text{RPI for month } x = &\text{RPI for previous January}\\ &\times (\text{all-item price ratio for month } x).\end{aligned}$$

The weights for a particular year are used in calculating the RPI for every month from February of that year to January of the following year.

Activity 21 Calculating the RPI for July 2011

Find the value of the RPI in July 2011 by completing the following table and the formulas below. The value of the RPI in January 2011 was 229.0. (The base date was January 1987.)

Table 14 Calculating the RPI for July 2011

Group	Price ratio for July 2011 relative to January 2011 r	2011 weights w	Price ratio × weight rw
Food and catering	1.024	165	
Alcohol and tobacco	1.042	88	
Housing and household expenditure	1.012	408	
Personal expenditure	1.053	82	
Travel and leisure	1.030	257	
Sum			

(Source: Office for National Statistics)

sum $(w) =$ ___, sum of products $(rw) =$ ___,

all-item price ratio $= \dfrac{\text{sum of products } (rw)}{\text{sum}(w)} =$ ___, value of RPI in July 2011 $=$ ___.

The published value for the RPI in July 2011 was 234.7, slightly different from the value you should have obtained in Activity 21 (that is, 234.6). The discrepancy arises because the government statisticians use more accuracy during their RPI calculations, and round only at the end before publishing the results.

The following activity is intended to help you draw together many of the ideas you have met in this section, both about what the RPI is and how it is calculated.

Activity 22 The effects of particular price changes on the RPI

Between February 2011 and February 2012, the price of leisure goods fell on average by 2.3%, while the price of canteen meals rose by 2.8%. Answer the following questions about the likely effects of these changes on the value of the RPI. (No calculations are required.)

(a) Looked at in isolation (that is, supposing that no other prices changed), would the change in the price of leisure goods lead to an increase or a decrease in the value of the RPI?

Would the change in the price of canteen meals (looked at in isolation) lead to an increase or a decrease in the value of the RPI?

(b) In each case, is the size of the increase or decrease likely to be large or small?

(c) Using what you know about the structure of the RPI, decide which of ‘Leisure goods’ and ‘Canteen meals’ has the larger weight.

(d) Which of the price changes mentioned in the question will have a larger effect on the value of the RPI? Briefly explain your answer.

5.3 Using the price indices

The RPI and CPI are intended to help measure price changes, so we shall start this section by describing how to use them for this purpose.

Example 23 A news report on inflation

The BBC News website reported (20 March 2012) ‘UK inflation rate falls to 3.4% in February’. What does that actually mean?

The rest of the BBC article makes it clear that this ‘inflation’ figure was based on the CPI rather than the RPI, but its meaning is still not obvious. What is usually meant in situations like this is the following.

The annual rate of inflation

In the UK, the (annual) rate of inflation is the percentage increase in the value of the CPI (or the RPI) compared to one year earlier.

(In M140, it will always be made clear whether you should use the CPI or the RPI in contexts like this.)

The annual rate of inflation is sometimes called the *year-on-year rate of inflation*.

In February 2012, the CPI was 121.8. Exactly a year earlier, in February 2011, the CPI was 117.8. The ratio of these two values is

$$\frac{\text{value of CPI in February 2012}}{\text{value of CPI in February 2011}} = \frac{121.8}{117.8} \simeq 1.034.$$

So the value of the CPI in February 2012 was 3.4% higher than in the previous February. That is the source of the number in the BBC headline.

Activity 23 The annual inflation rate in February 2012

In February 2012, the RPI was 239.9. Exactly a year earlier, in February 2011, the RPI was 231.3. Calculate the annual inflation rate for February 2012, based on the RPI.

The fact that the inflation rates that are generally reported in the media relate to price *increases* (as measured in a price index) over a *whole year* means that one has to be careful in interpreting the figures, in several ways.

- Media reports might say that 'inflation is falling', but this does not mean that *prices* are falling. It simply means that the annual inflation rate is less than it was the previous month. So when the BBC headline said that the (annual) inflation rate had fallen to 3.4% in February 2012, it meant that the February 2012 rate was smaller than the January 2012 rate (which was 3.6%). Prices were still rising, but not quite so quickly.
- The change in price levels over one month may be, and indeed usually is, considerably different from the annual inflation rate. For instance, prices actually fell between December 2011 and January 2012: the CPI was 121.7 in December 2011 and 121.1 in January 2012. (Prices in the UK usually fall between December and January in the UK, as Christmas shopping ends and

the January sales begin.) But the annual inflation rate for January 2012, measured by the CPI, was 3.6%.

- The effect of a single major cause of increased prices can persist in the annual inflation rates long after the prices originally increased. For instance, the standard rate of value added tax (VAT) in the UK went up from 17.5% to 20% at the start of January 2011, causing a one-off increase in the price (to consumers) of many goods and services. This showed up in the annual inflation rate for January 2011, where prices were 4.0% higher than a year earlier. Moreover, the annual inflation rate for every other month in 2011 was also affected by the VAT increase, because in each case the CPI was being compared to the CPI in the corresponding month in 2010, before the VAT increase.

Another important use of price indices like the RPI and CPI is for *index-linking*. This is used for such things as savings and pensions, as a means of safeguarding the value of money held or received in these forms.

Index-linking an amount

To index-link any amount of money, the amount in question is multiplied by the same ratio as the change in the value of the price index. Another term for this process is **indexation**.

It is important to stress the notion of *ratio* in index-linking, because it is only by calculating the ratio of two indices that you can get an accurate measure of how prices have increased. For example, an increase in the RPI from 100 to 200 represents a 100% increase in price, whereas a further RPI increase from 200 to 300 represents only a further 50% increase in price.

Example 24 Index-linking a pension

The value of the RPI for February 2012 was 239.9 whereas the corresponding figure for February 2011 was 231.3. So an index-linked pension that was, say, \$450 per month in February 2011, would be increased to

$$\$450 \times \frac{239.9}{231.3} \text{ (i.e. \$466.73) per month}$$

for February 2012. The reason for index-linking the pension in this way is that the increased pension would buy the same amount of goods or services in February 2012 as the original pension bought in February 2011 – that is, it should have the same purchasing power.

Pensions can be, and indeed increasingly are, index-linked using the CPI rather than the RPI.

Activity 24 Index-linking a pension using the CPI

An index-linked pension was \$120 per week in November 2010. It is index-linked using the CPI. How much should the pension be per week in November 2011? The value of the CPI was 115.6 in November 2010 and 121.2 in November 2011.

This principle leads to another much-quoted figure which can be calculated directly from the RPI: **the purchasing power of the pound**. (This is the purchasing power of the pound *within this country*, not its purchasing power abroad; the latter is a distinct and far more complicated concept.) The purchasing power of the pound measures how much a consumer can buy with a fixed amount of money at one point of time compared with another point of time.

The word *compared* here is again important; it makes sense only to talk about the purchasing power of the pound at one time *compared* with another. For example, if \$1 worth of goods would have cost only 60p four years ago, then we say that the purchasing power of the pound is only 60p compared with four years earlier.

Purchasing power of the pound

The purchasing power (in pence) of the pound at date A compared with date B is

$$\frac{\text{value of RPI at date B}}{\text{value of RPI at date A}} \times 100.$$

The purchasing power of the pound could be calculated using the CPI instead, though the figures published by the Office for National Statistics do happen to use the RPI.

Example 25 Calculating the purchasing power of the pound

(a) The purchasing power of the pound in February 2012 compared with February 2011 was

$$\frac{231.3}{239.9} \times 100\text{p} = 96.415\,17\text{p}.$$

231.3 and 239.9 are the two RPI values given in Activity 23.

We round this to give 96p.

(b) The purchasing power of the pound in February 2012 compared with the base date, January 1987, was

$$\frac{100}{239.9} \times 100\text{p}.$$

(At the base date, the value of the RPI is 100 by definition.)

This is, after rounding, 42p.

Activity 25 Annual inflation and the purchasing power of the pound

Table 15 Values of the RPI from January 2009 to December 2011

Month	2009	2010	2011	Month	2009	2010	2011
January	210.1	217.9	229.0	July	213.4	223.6	234.7
February	211.4	219.2	231.3	August	214.4	224.5	236.1
March	211.3	220.7	232.5	September	215.3	225.3	237.9
April	211.5	222.8	234.4	October	216.0	225.8	238.0
May	212.8	223.6	235.2	November	216.6	226.8	238.5
June	213.4	224.1	235.2	December	218.0	228.4	239.4

(Source: Office for National Statistics)

For each of the following months, use the values of the RPI in Table 15 to calculate the annual inflation rate (based on the RPI) and to calculate the purchasing power of the pound (in pence) compared to one year previously.

(a) May 2010 (b) October 2011 (c) March 2011

You have seen that the RPI can be used as a way of updating the value of a pension to take account of general increases in prices (index-linking). The RPI is used in other similar ways, for instance to update the levels of some other state benefits and investments. But the CPI *could* be used for these purposes.

Why are there two different indices? Let's look at how this arose. As well as its use for index-linking, which is basically to compensate for price changes, the RPI previously played an important role in the management of the UK economy generally. The government sets targets for the rate of inflation, and the Bank of England Monetary Policy Committee adjusts interest rates to try to achieve these targets. Until the end of 2003, these inflation targets were based on the RPI, or to be precise, on another price index called RPIX which is similar to the RPI but omits owner-occupiers' mortgage interest payments from the calculations. (There are good economic reasons for this omission, to do with the fact that in many ways the purchase of a house has the character of a long-term investment, unlike the purchase of, say, a bag of potatoes.) From 2004, the inflation targets have instead been set in terms of the CPI. The CPI is calculated in a way that matches similar inflation measures in other countries of the European Union. (So it can be used for international comparisons.)

In terms of general principles, though, and also in terms of most of the details of how the indices are calculated, the differences between the RPI and CPI are not actually very great. As mentioned in Subsection 5.1, the CPI reflects the spending of a wider population than the RPI. Partly because of this, there are certain items (e.g. university accommodation fees) that are included in the CPI but not the RPI. There are also certain items that are included in the RPI but not the CPI, notably some owner-occupiers' housing costs such as mortgage interest payments and house-building insurance. Finally, the CPI uses a different method to the RPI for combining individual price measurements.

Because of these differences, inflation as measured by the CPI tends usually to be rather lower than that measured by the RPI. In Example 23, you saw that the annual inflation rate in February 2012 as measured by the CPI was 3.4%. The annual inflation rate in the same month, as measured by the RPI, was 3.7%, as you saw in Activity 23. The RPI continues to be calculated and published, and to be used to index-link payments such as savings rates and some pensions. However, there are reasons why the RPI is more appropriate than the CPI for *some* such purposes, and it seems likely to continue in use for a long time. Furthermore, changes in how index-linking is done can be politically very controversial. For instance, in 2010, the UK government announced that in future, public sector pensions would be index-linked to the CPI rather than the RPI, which caused major complaints from those affected (because inflation as measured by the CPI is usually lower than that measured using the RPI, so pensions will not increase so much in money terms).

Arguably it is rather strange to use the RPI to index pensions, given that (as was said at the beginning of Subsection 5.1) the RPI omits the expenditure of pensioner households.

You might be asking yourself which is the 'correct' measure of inflation – RPI, CPI, or something else entirely. There is no such thing as a single 'correct' measure. Different measures are appropriate for different purposes. That's why it is important to understand just what is being measured and how.

In this section, you have seen how price rises are measured using an index of retail prices. Earnings are discussed in the next unit. Only when prices and earnings have both been considered can you begin to answer the central question of these two units: *Are people getting better or worse off?* In the next unit, you will see how to use a price index in conjunction with an index of earnings to see whether rises in earnings are keeping pace with rises in prices.

Exercises on Section 5

Exercise 10 Calculating the RPI for February 2012

Find the value of the RPI in February 2012, using the data in the table below. The value of the RPI in January 2012 was 238.0.

Table 16 Calculating the RPI for February 2012

Group	Price ratio for February 2012 relative to January 2012 r	2012 weights w	Price ratio × weight rw
Food and catering	1.009	161	
Alcohol and tobacco	1.005	85	
Housing and household expenditure	1.003	412	
Personal expenditure	1.040	84	
Travel and leisure	1.005	258	
Total			

(Source: Office for National Statistics)

Exercise 11 Annual inflation rates and the purchasing power of the pound

For each of the following months, use Table 15 (in Subsection 5.3) to calculate the annual inflation rate given by the RPI and to calculate the purchasing power of the pound (in pence) compared to one year previously.

(a) October 2010

(b) January 2011

Exercise 12 Index-linking another pension

An index-linked pension (linked to the RPI) was £800 per month in April 2010. How much should it be in April 2011? (Again, use the RPI values in Table 15.)

6 Computer work: measures of location

In Subsection 1.4, you learned that the median is a resistant measure and the mean is a sensitive measure. You will explore what this means in practice for a particular dataset and then verify the rules for weighted means for a particular example. You should work through all of Chapter 2 of the Computer Book now, if you have not already done so.

Summary

In this unit you have been discovering how statistics can be used to answer questions about prices. You have learned how to find a single number to summarise the price of an item at a particular point in time, even though the item might be available from a number of sources. You have also learned how to combine information on prices across a range of goods and services. Then, through the use of price ratios, you have seen how changes in price over time can be quantified. In particular, you have learned about chained price indices such as the Retail Prices Index (RPI) and Consumer Prices Index (CPI), used in the UK to measure inflation.

Two more measures of location, the mean and weighted mean, have been introduced. The mean is a *sensitive* measure whereas the median is a *resistant* measure. The weighted mean only depends on the relative sizes of the weights, and the weighted mean of two numbers is always closer to the value with the highest weight.

You have learned about measures of spread, in particular the range and the interquartile range, and about quartiles, from which the interquartile range is calculated. The five-figure summary was described, which consists of the minimum, lower quartile, median, upper quartile and maximum, along with the size of the batch. A way of displaying the five-figure summary, the boxplot, was introduced. The 'box' in the boxplot runs between the lower and upper quartiles and has a line in it corresponding to the median, thus displaying three of the five numbers in the five-number summary. The other two numbers in the five-number summary, the minimum and maximum, are given by the lengths of the whiskers or position of potential outliers.

You have learned how the RPI and the CPI are calculated by the Office for National Statistics from a 'basket' of goods using weighted means to give price ratios, group price ratios and all-commodities price ratios. These all-commodity price ratios are then chained to give the value of the index relative to a base date. The RPI and CPI can be used to calculate inflation, to index-link amounts of money and to calculate the purchasing power of the pound at one time compared with another.

Learning outcomes

After working through this unit, you should be able to:

- find the median of a batch of data
- find the mean of a batch of data
- describe what is meant by a resistant measure of location, and identify which measures are resistant
- find the weighted mean of two numbers with associated weights
- use the weighted mean to combine two batch means to find the mean of the combined batch
- use the weighted mean to find the overall average cost of a commodity from the price paid and quantity purchased on two occasions
- understand the use of a weighted mean in other contexts and for larger sets of numbers
- find the upper and lower quartiles and the interquartile range of a batch of data
- prepare a five-figure summary of a batch of data
- interpret the boxplot of a batch of data
- use the boxplot to investigate the overall shape of a batch of data, in particular its symmetry and skewness
- calculate a simple chained price index and explain what is meant by its base date
- describe the major steps in producing the Retail Prices Index
- calculate the value of the Retail Prices Index from the five group price ratios and weights
- use the Retail Prices Index or the Consumer Prices Index to compare the general level of prices at two dates and calculate the rise in the general level of prices over a year (the annual rate of inflation)
- use the Retail Prices Index or the Consumer Prices Index to do index-linking calculations, and use the Retail Prices Index to find the purchasing power of the pound at one date compared with another.

Solutions to activities

Solution to Activity 1

For a batch size of 20, the median position is $\frac{1}{2}(20+1) = 10\frac{1}{2}$. So, the median will be halfway between $x_{(10)}$ and $x_{(11)}$. These are both 150, so the median is $150.

Solution to Activity 2

(a) A stemplot of all 14 prices in the table is shown below.

374	0 0 3
375	
376	0 7
377	6
378	4
379	5 6
380	1 1 4 5
381	8

$n = 14$ 374 | 0 represents 3.740p per kWh

Stemplot of 14 gas prices

(b) Stemplots for the prices for northern and southern cities are shown below.

Northern

374	0 0
375	
376	7
377	6
378	
379	
380	1 1 4

$n = 7$ 374 | 0 represents 3.740p per kWh

Southern

374	3
375	
376	0
377	
378	4
379	5 6
380	5
381	8

$n = 7$ 374 | 3 represents 3.743p per kWh

Stemplots for northern and southern cities separately.

(c) For a batch size of 14, the median position is $\frac{1}{2}(14+1) = 7\frac{1}{2}$. So, the all-cities median will be halfway between $x_{(7)}$ and $x_{(8)}$. These are 3.784 and 3.795, so the median is 3.7895, which is 3.790 when rounded to three decimal places. (The rounded median should be written as 3.790 and not 3.79, to show it is accurate to three decimal places and not just two.)

For the northern and southern batches, both of size 7, the median for each is the value of $x_{(4)}$ (that is, $\frac{1}{2}(7+1) = 4$). This is 3.776 for the northern batch and 3.795 for the southern batch.

The range is the difference between the upper extreme, E_U, and the lower extreme, E_L (range $= E_U - E_L$). So the all-cities range is

$$3.818 - 3.740 = 0.078,$$

the range for the northern batch is

$$3.804 - 3.740 = 0.064,$$

and the range for the southern batch is

$$3.818 - 3.743 = 0.075.$$

The medians and ranges are summarised below.

	Median	Range
All cities	3.790	0.078
Northern cities	3.776	0.064
Southern cities	3.795	0.075

Thus the general level of gas prices in the country as a whole was about 3.790p per kWh. The average price differed by only 0.078p per kWh across the 14 cities.

The difference between the median prices for the northern and southern cities is 0.019p per kWh $(3.795 - 3.776 = 0.019)$, with the south having the higher median.

The analysis does not clearly reveal whether the general level of gas prices for typical consumers in 2010 was higher in the south or in the north, though there is an indication that prices were a little higher in the south. The range of prices was also rather greater in the south. It is worth noting that the differences in gas prices between the cities in Table 3 were generally small, when measured in pence per kWh – although, with a typical annual gas usage of 18 000 kWh, the price difference between the most expensive city and the cheapest would amount to an annual difference in bills of about \$14 on a typical bill of somewhere around \$700.

Solution to Activity 3

Using the data for the prices from Activity 1:

$$\text{mean} = \frac{\text{sum}}{\text{size}} = \frac{90 + 100 + \ldots + 270}{20} = \$162.$$

Or using the $\sum$ notation, $\sum x = 90 + 100 + \ldots + 270 = 3240$ and $n = 20$, so

$$\text{mean} = \overline{x} = \frac{\sum x}{n} = \frac{3240}{20} = \$162.$$

The prices were rounded to the nearest \$10, so it is appropriate to keep one more significant figure for the mean, that is, to show it accurate to the nearest \$1. So since the exact value is \$162, it needs no further rounding.

Solution to Activity 4

The entries are

Mean	Median
3.7859	3.795
3.7996	3.796

Whereas deletion of Cardiff and Ipswich has the effect of increasing the mean price by 0.0137p per kWh, the median price increases by only 0.001p per kWh. This is what we would expect as, in general, the more resistant a measure is, the less it changes when a few extreme values are deleted.

Solution to Activity 5

The entries are

Mean	Median
3.7996	3.796
4.6996	3.796

Here the median is completely unaffected by the misprint, although the mean changes considerably.

Solution to Activity 6

You should expect the weighted mean price to be nearer the London price, because of Rule 2 for weighted means (Subsection 2.1) and given that London has a much larger weight then Edinburgh.

The weighted mean price given by the formula in Example 11 is (after rounding) 3.814p per kWh, which is indeed much closer to the London price than to the Edinburgh price.

Solution to Activity 7

(a) OCAS $= \dfrac{(80 \times 50) + (60 \times 50)}{50 + 50} = \dfrac{4000 + 3000}{100} = \dfrac{7000}{100} = 70.$

This is the same as a simple (unweighted) mean of the two scores, because the two component scores have equal weight. It lies exactly halfway between the two scores ($\frac{1}{2}(80 + 60) = 70$).

(b) OCAS $= \dfrac{(80 \times 40) + (60 \times 60)}{40 + 60} = \dfrac{3200 + 3600}{100} = \dfrac{6800}{100} = 68.$

This is slightly less than the simple mean in (a) because the component with the lower score (TMA) has the greater weight.

(c) OCAS $= \dfrac{(80 \times 65) + (60 \times 55)}{65 + 55} = \dfrac{5200 + 3300}{120} = \dfrac{8500}{120} \simeq 70.8.$

This is slightly higher than the simple mean in (a) because the component with the higher score (iCMA) has the greater weight.

(Note that the weights need not necessarily sum to 100, even when dealing with percentages.)

(d) OCAS $= \dfrac{(80 \times 25) + (60 \times 75)}{25 + 75} = \dfrac{2000 + 4500}{100} = \dfrac{6500}{100} = 65.$

This is even lower than (b), so even nearer the lower score (TMA), because the TMA score has even greater weight.

(e) OCAS $= \dfrac{(80 \times 30) + (60 \times 90)}{30 + 90} = \dfrac{2400 + 5400}{120} = \dfrac{7800}{120} = 65.$

This is the same as (d) because the ratios of the weights are the same; they are both in the ratio 1 to 3. That is, $25 : 75 = 30 : 90$ $(= 1 : 3)$.

(We say this as follows: ‘the ratio 25 to 75 equals the ratio 30 to 90’.)

Solution to Activity 9

The table showing the required sums (and the values in the xw column, that you may not have had to write down), is as follows.

	Price (p/kWh) x	Weight w	Price × weight xw
Aberdeen	13.76	19	261.44
Belfast	15.03	58	871.74
Edinburgh	13.86	42	582.12
Leeds	12.70	150	1 905.00
Liverpool	13.89	82	1 138.98
Manchester	12.65	224	2 833.60
Newcastle-upon-Tyne	12.97	88	1 141.36
Nottingham	12.64	67	846.88
Birmingham	12.89	228	2 938.92
Canterbury	12.92	5	64.60
Cardiff	13.83	33	456.39
Ipswich	12.84	14	179.76
London	13.17	828	10 904.76
Plymouth	13.61	24	326.64
Southampton	13.41	30	402.30
Sum		1892	24 854.49

Thus $\sum xw = 24\,854.49$, $\sum w = 1892$ and

$$\frac{\sum xw}{\sum w} = \frac{24\,854.49}{1892} = 13.136\,623 \simeq 13.14.$$

So the weighted mean of electricity prices is 13.14p per kWh.

Solution to Activity 10

(a) Here, because $n = 15$, an appropriate picture of the data would be Figure 9. To find the lower and upper quartiles, Q_1 and Q_3, of this batch, first find $\frac{1}{4}(n+1) = 4$ and $\frac{3}{4}(n+1) = 12$. Therefore $Q_1 = 268$p and $Q_3 = 299$p.

(b) For this batch, $n = 14$ so $\frac{1}{4}(n+1) = 3\frac{3}{4}$ and $\frac{3}{4}(n+1) = 11\frac{1}{4}$.

$$\begin{aligned} Q_1 &= 3.743 + \tfrac{3}{4}(3.760 - 3.743) \\ &= 3.755\,75 \simeq 3.756 \end{aligned}$$

and

$$\begin{aligned} Q_3 &= 3.801 + \tfrac{1}{4}(3.804 - 3.801) \\ &= 3.801\,75 \simeq 3.802. \end{aligned}$$

So the lower quartile is 3.756 p per kWh and the upper quartile is 3.802p per kWh.

Solution to Activity 11

The range is the distance between the extremes:

$$\begin{aligned} \text{range} &= E_U - E_L \\ &= 369\text{p} - 268\text{p} \\ &= 101\text{p}. \end{aligned}$$

The interquartile range is the distance between the quartiles:

$$\begin{aligned} \text{IQR} &= Q_3 - Q_1 \\ &= 299\text{p} - 268\text{p} \\ &= 31\text{p}. \end{aligned}$$

Solution to Activity 12

The quartiles, before rounding, are $Q_1 = 3.755\,75$ and $Q_3 = 3.801\,75$. So

$$\begin{aligned}\text{IQR} &= Q_3 - Q_1\\ &= 3.801\,75 - 3.755\,75\\ &= 0.046,\end{aligned}$$

and the interquartile range is 0.046p per kWh.

Solution to Activity 13

(a) All the necessary figures have already been calculated. You found the median (3.790) in Activity 2 and the quartiles ($Q_1 = 3.756$, $Q_3 = 3.802$) in Activity 10. The extremes ($E_L = 3.740$, $E_U = 3.818$) and the batch size ($n = 14$) are clearly shown in the stemplot.

So the five-figure summary is as follows:

		3.790	
$n = 14$	3.756		3.802
	3.740		3.818

(b) Looking at the stemplot, on the whole the lower values are more spread out, indicating that the data are not symmetric and are left-skew.

The central box of the boxplot again shows left skewness, with the left-hand part of the box being clearly longer than the right-hand part. However, this skewness does not show up in the lengths of the whiskers in this batch – they are both the same length.

Solution to Activity 14

The increase (in \$/MWh) is $29 - 24 = 5$. This is $\frac{5}{24} \simeq 0.208$ as a proportion of the 2007 price. That is, $\frac{5}{24} \times 100\% \simeq 20.8\%$ of the 2007 price. Or you might have worked this out by finding that the 2008 price is $\frac{29}{24} \times 100\% \simeq 120.8\%$ of the 2007 price, so that again the increase is 20.8% of the 2007 price.

Solution to Activity 15

The 2008 electricity price is $1.145 \times 100\% = 114.5\%$ of the 2007 price, so that the increase is 14.5% of the 2007 price.

The 2008 value of the electricity price index is

(value of the index in 2007, which is 100)

$\times$ (electricity price ratio for 2008 relative to 2007)

$= 100 \times 1.145 = 114.5.$

Solution to Activity 16

The expenditure on a particular fuel in a particular year can be calculated as expenditure $=$ quantity used $\times$ price. Therefore, if the expenditure and price are known, the quantity used can be calculated as

$$\text{quantity used} = \frac{\text{expenditure}}{\text{price}}.$$

In 2007, Gradgrind's gas cost \$24 per MWh, and they spent \$9298 on gas, so the amount of gas they used in MWh was

$$\frac{9298}{24} \simeq 387.4.$$

The other amounts, in MWh, are found in a similar way, and all are shown in the following table.

	2007	2008
Gas	387.4	280.9
Electricity	42.2	34.4

The reason that the expenditures went down is simply that Gradgrind used less of each fuel in 2008 than in 2007.

Solution to Activity 17

(a) The gas price ratio for 2009 relative to 2008 is

$$\frac{30}{29} \simeq 1.034.$$

The electricity price ratio for 2009 relative to 2008 is

$$\frac{98}{87} \simeq 1.126.$$

(Over this year, electricity prices rose a lot more than gas prices.)

(b) The overall energy price ratio for 2009 relative to 2008 is

$$\frac{(1.034 \times 8145) + (1.126 \times 2991)}{8145 + 2991} = \frac{11\,789.796}{11\,136} \simeq 1.059.$$

(c) Using the 2009 expenditures for weights instead of the 2008 expenditures, the overall energy price ratio for 2009 relative to 2008 is

$$\frac{(1.034 \times 23\,733) + (1.126 \times 2275)}{23\,733 + 2275} = \frac{27\,101.572}{26\,008} \simeq 1.042.$$

This price ratio is considerably less than the one found in part (b).

Solution to Activity 18

The gas price ratio for 2010 relative to 2009 is

$$\frac{28}{30} \simeq 0.933.$$

The electricity price ratio for 2010 relative to 2009 is

$$\frac{88}{98} \simeq 0.898.$$

(Both price ratios are less than 1 because, over this year, Gradgrind's gas and electricity prices both fell.)

The overall energy price ratio for 2010 relative to 2009 is

$$\frac{(0.933 \times 23\,733) + (0.898 \times 2275)}{23\,733 + 2275} = \frac{24\,185.839}{26\,008} \simeq 0.930.$$

Then the value of the index for 2010 is found by multiplying the 2009 value of the index by this overall price ratio, giving

$$126.2 \times 0.930 \simeq 117.4.$$

Solution to Activity 19

(a) What you need to remember here is that the size of an area represents the proportion of expenditure on that class of goods or services. (Also, it is admittedly not very easy to estimate these areas 'by eye'! Your estimates might quite reasonably differ from those given here.)

- The sector for 'Personal expenditure' looks as if it is approximately a tenth of the whole inner circle – so approximately a tenth of total expenditure is personal expenditure.
- 'Housing and household expenditure' looks as if it is somewhere between a third and a half of the inner circle – perhaps approximately two fifths – so approximately two fifths of expenditure is on housing and household expenditure.
- The area for 'Housing' takes up about a quarter of the outer ring, so about a quarter of expenditure is on housing.

(b) The amount spent each week on 'Personal expenditure' is approximately

$$\frac{1}{10} \times \$540 = \$54.$$

The amount spent each week on 'Housing and household expenditure' is approximately

$$\frac{2}{5} \times \$540 = \$216 \simeq \$220.$$

The amount spent each week on 'Housing' is approximately

$$\frac{1}{4} \times \$540 = \$135 \simeq \$140.$$

Recall, however, that the weights represent *average* proportions of expenditure, and the spending patterns of the selected household may differ from those of the 'typical' household.

Solution to Activity 20

Every household will be different, but think about the reasons for any large differences between your weights and those for the RPI.

Solution to Activity 21

Group	Price ratio for July 2011 relative to January 2011 r	2011 weights w	Price ratio $\times$ weight rw
Food and catering	1.024	165	168.960
Alcohol and tobacco	1.042	88	91.696
Housing and household expenditure	1.012	408	412.896
Personal expenditure	1.053	82	86.346
Travel and leisure	1.030	257	264.710
Sum		1000	1024.608

$$\text{sum } (w) = 1000, \quad \text{sum of products } (rw) = 1024.608,$$

$$\begin{aligned}\text{all-item price ratio} &= \frac{\text{sum of products } (rw)}{\text{sum}(w)} = \frac{1024.608}{1000}\\ &= 1.024\,608,\end{aligned}$$

$$\begin{aligned}\text{value of RPI in July 2011} &= 229.0 \times 1.024\,608\\ &= 234.635\,232\\ &\simeq 234.6.\end{aligned}$$

Solution to Activity 22

More detail has been included in these comments than is expected from you. When you read them, make sure you understand all the points mentioned.

(a) The RPI is calculated using the price ratio and weight of each item. Since the weights of items change very little from one year to the next, the price ratio alone will normally tell you whether a change in price is likely to lead to an increase or a decrease in the value of the RPI. If a price rises, then the price ratio is greater than one, so the RPI is likely to increase as a result. If a price falls, then the price ratio is less than one, so the RPI is likely to decrease. Therefore, since the price of leisure goods fell, this is likely to lead to a decrease in the value of the RPI. For a similar reason, the increase in the price of canteen meals is likely to lead to an increase in the value of the RPI.

(b) Both changes are likely to be small for two reasons. First, the price changes are themselves fairly small. Second, leisure goods and canteen meals form only part of a household's expenditure: no single group, subgroup or section will have a large effect on the RPI on its own, unless there is a very large change in its price.

(c) The weight of 'Leisure goods' was 33 in 2012 (see Table 12). Since 'Canteen meals' is only one section in the subgroup 'Catering', which had weight 47 in 2012, the weight of 'Canteen meals' will be much smaller than 47. (In fact it was 3.) So the weight of 'Leisure goods' is much larger than the weight of 'Canteen meals'.

(d) Since the weight of 'Leisure goods' is much larger than the weight of 'Canteen meals', and the percentage change in the prices are not too different in size, the change in the price of leisure goods is likely to have a much larger effect on the value of the RPI as a whole.

Solution to Activity 23

The ratio of the two RPI values is

$$\frac{\text{value of RPI in February 2012}}{\text{value of RPI in February 2011}} = \frac{239.9}{231.3} \simeq 1.037,$$

or 103.7%. Therefore the annual inflation rate, based on the RPI was 3.7%. (Note that this is slightly higher than the annual inflation rate measured using the CPI.)

Solution to Activity 24

The weekly amount in November 2011 should be

$$\$120 \times \frac{121.2}{115.6} \simeq \$125.81.$$

Solution to Activity 25

(a) For May 2010, the ratio of the value of the RPI to its value one year earlier is

$$\frac{223.6}{212.8} \simeq 1.051,$$

so the annual inflation rate is 5.1%.

The purchasing power of the pound compared to one year previously is

$$\frac{212.8}{223.6} \times 100\text{p} \simeq 95\text{p}.$$

(b) For October 2011, the ratio of the value of the RPI to its value one year earlier is

$$\frac{238.0}{225.8} \simeq 1.054,$$

so the annual inflation rate is 5.4%.

The purchasing power of the pound compared to one year previously is

$$\frac{225.8}{238.0} \times 100\text{p} \simeq 95\text{p}.$$

(c) For March 2011, the ratio of the value of the RPI to its value one year earlier is

$$\frac{232.5}{220.7} \simeq 1.053,$$

so the annual inflation rate is 5.3%.

The purchasing power of the pound compared to one year previously is

$$\frac{220.7}{232.5} \times 100\text{p} \simeq 95\text{p}.$$

Solutions to exercises

Solution to Exercise 1

(a) For the arithmetic scores, the position of the median is $\frac{1}{2}(33+1) = 17$, so the median is 79%.

(b) For the television prices, the position of the median is $\frac{1}{2}(26+1) = 13\frac{1}{2}$, so the median is halfway between $x_{(13)}$ and $x_{(14)}$. Thus, the median is

$$\tfrac{1}{2}(\$269 + \$270) = \$269.5 \simeq \$270.$$

Solution to Exercise 2

For the batch of arithmetic scores in part (a) of Exercise 1, the sum of the 33 values is 2326 and

$$\frac{2326}{33} \simeq 70.5.$$

Therefore, the mean is 70.5%. (The original data are given to the nearest whole number, so the mean is rounded to one decimal place.)

For the batch of television prices in part (b) of Exercise 1, the sum of the 26 values is 7856 and

$$\frac{7856}{26} = 302.1538 \simeq 302.2.$$

Therefore, the mean is $302.2.

Solution to Exercise 3

For the median, there are now 17 prices left in the batch, so the median is at position $\frac{1}{2}(17+1) = 9$. It is therefore 150.

The sum of the remaining 17 values is 2480, so the mean is

$$\frac{2480}{17} = 145.8824 \simeq 145.9.$$

In this case, removing the three highest prices has not changed the median at all, but it has reduced the mean considerably. This illustrates that the median is a more resistant measure than the mean.

Solution to Exercise 4

Mean price of all the cameras is

$$\frac{(80.7 \times 10) + (78.5 \times 17)}{10 + 17} = \frac{2141.5}{27},$$

which is $79.3 (rounded to the same accuracy as the original means).

Solution to Exercise 5

Mean price of all the material is

$$\frac{(10.95 \times 8.5) + (12.70 \times 6)}{8.5 + 6} = \frac{169.275}{14.5},$$

which is $11.67 (rounded to the nearest penny).

Solution to Exercise 6

(a) For the arithmetic scores, $n = 33$ so $\frac{1}{4}(n+1) = 8\frac{1}{2}$ and $\frac{3}{4}(n+1) = 25\frac{1}{2}$.

The lower quartile is therefore

$$Q_1 = \tfrac{1}{2}(55 + 58)\% = 56.5\% \simeq 57\%.$$

The upper quartile is

$$Q_3 = \tfrac{1}{2}(86 + 89)\% = 87.5\% \simeq 88\%.$$

The interquartile range is

$$Q_3 - Q_1 = 87.5\% - 56.5\% = 31\%.$$

(b) For the television prices, $n = 26$ so $\frac{1}{4}(n+1) = 6\frac{3}{4}$ and $\frac{3}{4}(n+1) = 20\frac{1}{4}$.

The lower quartile is therefore

$$Q_1 = \$229 + \tfrac{3}{4}(\$230 - \$229) = \$229.75 \simeq \$230.$$

The upper quartile is

$$Q_3 = \$320 + \tfrac{1}{4}(\$349 - \$320) = \$327.25 \simeq £327.$$

The interquartile range is

$$Q_3 - Q_1 = \$327.25 - \$229.75 = \$97.5 \simeq \$98.$$

Solution to Exercise 7

(a) Arithmetic scores:

From the stemplot, $n = 33$, $E_L = 7$ and $E_U = 100$.

		79	
$n = 33$	57		88
	7		100

Five-figure summary of arithmetic scores

(b) Television prices:

From the data table, $n = 26$, $E_L = 170$ and $E_U = 699$.

		270	
$n = 26$	230		327
	170		699

Five-figure summary of television prices

Solution to Exercise 8

For the boxplot of arithmetic scores, the left part of the box is longer than the right part, and the left whisker is also considerably longer than the right. This batch is left-skew (as was also found in Unit 1 (Activity 20, Subsection 5.2)).

For the boxplot of television prices, the right part of the box is rather longer than the left part. The right whisker is also rather longer than the left, and if one also takes into account the fact that two potential outliers have been marked, the top 25% of the data are clearly much more spread out than the bottom 25%. This batch is right-skew.

Solution to Exercise 9

The gas price ratio for 2011 relative to 2010 is

$$\frac{30}{28} \simeq 1.071.$$

The electricity price ratio for 2011 relative to 2010 is

$$\frac{86}{88} \simeq 0.977.$$

The overall energy price ratio for 2011 relative to 2010 is

$$\frac{(1.071 \times 23\,969) + (0.977 \times 2920)}{23\,969 + 2920} = \frac{28\,523.639}{26\,889} \simeq 1.061.$$

Then the value of the index for 2011 is found by multiplying the 2010 value of the index by this overall price ratio, giving

$$117.4 \times 1.061 \simeq 124.6.$$

Solution to Exercise 10

$$\sum w = 1000, \quad \sum rw = 1007.760,$$

$$\begin{aligned}\text{all-item price ratio} &= \frac{\sum rw}{\sum w} = \frac{1007.760}{1000} \\ &= 1.007\,760,\end{aligned}$$

$$\begin{aligned}\text{value of RPI in February 2012} &= 238.0 \times 1.007\,760 \\ &= 239.846\,88 \\ &\simeq 239.8.\end{aligned}$$

(The published index was 239.9. Again, the difference between this and your calculated value is because the ONS statisticians used more accuracy in their intermediate calculations.)

Solution to Exercise 11

(a) For October 2010, the ratio of the value of the RPI to its value one year earlier is

$$\frac{225.8}{216.0} \simeq 1.045,$$

so the annual inflation rate is 4.5%.

The purchasing power of the pound compared to one year previously is

$$\frac{216.0}{225.8} \times 100\text{p} \simeq 96\text{p}.$$

(b) For January 2011, the ratio of the value of the RPI to its value one year earlier is

$$\frac{229.0}{217.9} \simeq 1.051,$$

so the annual inflation rate is 5.1%.

The purchasing power of the pound compared to one year previously is

$$\frac{217.9}{229.0} \times 100\text{p} \simeq 95\text{p}.$$

Solution to Exercise 12

The RPI for April 2011 was 234.4 and the RPI for April 2010 was 222.8. So in April 2011, the pension should be

$$\$800 \times \frac{234.4}{222.8} \simeq \$842 \text{ per month.}$$

Acknowledgements

Grateful acknowledgement is made to the following sources:

Table 3 Adapted from: https://www.gov.uk/government/statistical-data-sets/annual-domestic-energy-price-statistics

Table 5 Taken from: http://en.wikipedia.org/wiki/List_of_conurbations_in_the_United_Kingdom. This file is licensed under the Creative Commons Attribution Licence http://creativecommons.org/licenses/by/3.0/

Table 6 Department of Energy and Climate Change

Tables 13–15 Office for National Statistics licensed under the Open Government Licence v.1.0

Table 16 Adapted from data from the Office for National Statistics licensed under the Open Government Licence v.1.0

Figure 28 Crown copyright material is reproduced under Class Licence Number C01W0000065 with the permission of the Controller, Office of Public Sector Information (OPSI)

Subsection 1.1 figure, ‘Data, data, data!’, Mary Evans Picture Library

Subsection 1.2 figure, ‘An upside down V-shape’, GIDZY / www.flickr.com. This file is licensed under the Creative Commons Attribution Licence http://creativecommons.org/licenses/by/3.0/

Subsection 1.2 figure, ‘Not that kind of flat screen’, Joey Gannon / www.flickr.com. This file is licensed under the Creative Commons Attribution-Share Alike Licence http://creativecommons.org/licenses/by-sa/3.0/

Subsection 3.2 figure, ‘More birds, now showing the shape of the /\/\ diagram’, JUMBERO / www.flickr.com. This file is licensed under the Creative Commons Attribution-Share Alike Licence http://creativecommons.org/licenses/by-sa/3.0/

Subsection 3.3 quote from McCullagh, P. (2003): The Royal Statistical Society

Subsection 3.3 photo of John Tukey: Taken from http://rchsbowman.wordpress.com/2011/09/03/statistics-notes-%E2%80%94-biography-%E2%80%94-john-wilder-tukey/

Subsection 3.3 cartoon: www.causeweb.org

Subsection 5.2 quote from BBC News website, 14 March 2012: Taken from www.bbc.co.uk/news/business-17356286

Every effort has been made to contact copyright holders. If any have been inadvertently overlooked the publishers will be pleased to make the necessary arrangements at the first opportunity.

Index